UNIVERSITY OF OULU

FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Youngsu Dong

3D Virtual World as an Enabler for a Hybrid Virtual-Physical Situated Civic Engagement Platform

Master’s Thesis
Degree Programme in Computer Science and Engineering
May 2017

Youngsu Dong. (2017) 3D Virtual World as an Enabler for a Hybrid Virtual-Physical Situated Civic Engagement Platform. University of Oulu, International Master's Programme in Computer Science and Engineering (Ubiquitous Computing). Master's Thesis, 72 p.

ABSTRACT

User participatory design concept in the domain of urban computing has been playing a significant role around the industry and research area since it first came out. Many researchers and organisations involved in this strive to reach out a larger diversity of people so they could build urban environment better supportive and salubrious towards the community members. In this regard, we created a common ground platform for sharing opinions of people within the society by delivering a free speech from the public place. The installation is evaluated in our study dedicated in the context of human psychological sensation systems in the virtual environment, tries to find the relation between the performance of the task (giving a speech) carried out in a virtual space and the user's degree of presence and immersion.

The results of our series of field experiments show that there is a positive association between the system user interface and the quality of work, though, we could not extend it to the statement which saying the superior performance is the direct result of high immersion and presence.

Keywords: urban computing, urban structure/installation, urban intervene, user participatory architecture, social computing, virtual reality, virtual space, virtual environment, presence, immersion, meshmoon, 3D city modeling, community, society, speech performance measure, web-based virtual application.

TABLE OF CONTENTS

TABLE OF FIGURES

FOREWORD

This thesis was completed at the Center for Ubiquitous Computing, University of Oulu. I would like to extend my sincere gratitude to Dr Hannu Kukka for the supervision of the work and valuable feedback in general. His expertise and knowledge in research and work domain field helped me to build the foundation for the future career.

I am also grateful to other researchers who were involved in the project, Dr Marko Jurmu, Dr Mina Pakanen, Paula Alavesa, and Toni Alatalo.

Lastly, I would like to thank my team members of the project, Jilin Yang and Kai Wang for the best teamwork and the collaboration; and all of my friends in Finland and family back in Korea who have been giving me an endless trust and love.

Oulu, 21.5.2017

Youngsu Dong

ABBREVIATIONS

API	Application programming interface
HCI	Human-computer interaction
LED	Light-emitting diode
p	Probability
PhD	Doctor of philosophy
r	correlation and coefficient
RPC	Remote procedure call
TXML	Tiny extensible markup language
UDP	User datagram protocol
XML	eXtensible markup language

1. INTRODUCTION

1.1. Motivation

This thesis describes a novel virtual-physical hybrid community participatory platform; an urban in-situ media architecture that enables on-/offline community members to participate in activities with the aim of content creation, decision making, and open debate. This thesis mainly focuses on the in-depth description of design, implementation, and evaluation of the Virtual Soapbox System (VSS). The system is motivated by existing design concepts such as urban computing, urban intervention, participatory design.

The Soapbox project began in June 2015 with the aim of creating a virtual-physical hybrid community platform to connect the community members and open up the communication between each other. The definition of word 'Soapbox' is well described at 'Merriam-Webster' as follow:

'An improvised platform used by a self-appointed, spontaneous, or informal orator; broadly: something that provides an outlet for delivering opinions' [1].

The Soapbox System (SS) allows users to join in a community discussion. Major challenge is to motivate the minority groups of the society who are often marginalised in terms of representativeness. Much research efforts in the past have been devoted to involving a large cross section of the population in the community [2] [3] [4]. The benefits of a successful engagement are remarkable, and will allow us to approach widely divergent opinions, ideas, stories, and experiences of the community.

As part of SS systems, Virtual Soapbox System (VSS) offers socially inactive members of the society a chance to readily participate the public activities without trepidation.

This thesis focuses on the VSS for the online users. To meet the requirements of the system, we rigorously follow the processes of design, implementation, and evaluation; the brief descriptions of each stage are as follows:

First, the design goal of VSS defines a set of requirements for the system. For instance, high quality and realistic visual displays are a prominent aspect of designing the environment. As noted by Mennecke et al. (2007) and Witmer & Singer (1998), building an engaging environment in a VE relies fundamentally on the environment. With that achieved, our brains can expect and accept inputs from the mediated space [5] [6].

Second, the goal for system implementation of VSS. VE is an open world in which online users can stroll around the synthesised space. Having said that, the system has a high chance of being compromised throughout its operation. We would spend much efforts and time to ensure the system to work as expected under any circumstance.

Lastly, the goals for evaluation phase. The system evaluation process needs to walk through iterative assessment along the lines of implementation stage. Since the team collaboration is essential as SS is modularized into three sub-systems, carrying out system evaluation would require a careful and repeated collaboration by the team members.

With all objectives defined here, we will steer into the project work step by step, throughout the following chapters.

1.2. Research questions and hypothesis

This thesis focuses on knowledge about computationally mediated environment, in which Real World (RW) tasks are performed. The research questions comprise of the following:

Q. How does sense of immersion and presence influence on the task performance in a VE?
   a. What is the relation between immersion and presence of the user and the task performance of giving a speech in a VE?
   b. What other variables possibly affect the speech performance?

And we define a hypothesis as:

H0. The immersion and presence of the user in a VE have no correlation with the task performance.

We are studying the sense of immersion and presence of the user in the aspect of their influence on task performance in a VE; in this particular research context, the task refers to the activity of delivering a speech to the public. We evaluate the performance through several methods, using both quantitative data (questionnaire) as well as qualitative data (interview).

Statistical analysis will be used to measure the relationship between the psychologic sensation of the user and the quality of task fulfilment in a VE.

We also collect system logs to complement the performance measure, such as comments, votes and so on.

A series of field experiments (semi-controlled) were carried out for data collection. The participants were asked to prepare a public speech lasting approximately 10 to 20 minutes. The measurement of immersion and presence levels relies on the scores of the dedicated questionnaire filled by the participants after the experiment session; it is taken from a literature devoted to the quantitative measurement of immersion and presence by Witmer & Singer [5] [7].

2. BACKGROUND

2.1. Technologically mediated civic engagement

SS is a civic engagement community platform that employs participatory design with the goal of encouraging community members to take part in public activities occurring in urban environments. Traditionally, media architectures operate in various design forms with different methodologies. There is a shared interest in them; they strive to encourage people to participate and produce useful content or share ideas. The voluntary behaviour of users can contribute to many common issues such as, urban decision making, and city planning [8].

The term participatory design has a relatively long history in the context of environment design and computer system rooted in Scandinavia [8]. Early practices of a consensus-based urban architecture were introduced by an Italian architect Giancarlo De Carlo, who was an evangelist towards the concept of 'participatory architecture' [9]. Soon after it emerged, the concept has been broadly employed in architecture, urban planning, and many other relevant domains over the early years. And with it, rapid development in the computer science and technology has proposed a new way of human-computer interaction, which sooner or later becomes one of the most important approaches for the urban planning scenarios [10] [11] [12]. Similarly, over the last decade, advancing technology in commodity displays together with the emergence of new interaction modalities such as gestures and eye movements have enabled a high potential to the future ubiquitous systems situated in urban environments [8].

Throughout the following chapters, we will introduce some of the existing practices developed with the community participatory design. These urban situated architectures also share other design concepts such as community consultation, community memory, urban intervention, and urban/ubiquitous computing.

2.1.1. Early practices of social interaction involving public displays

Early in 1980s, the first move toward connecting two physically remote places by using displays emerged. The work took the form of 'media links', using video and audio links to connect New York City and 'The Broadway' department store in Los Angles. This experiment, titled as 'Hole-In-Space' was carried out by Kit Galloway and Sherrie Rabinowitz for three days in November 1980. The installation exchanged video feed of both places, which allowed people in New York City communicate with people in L.A. real-time [13] [14].

Later projects in this period experimented with longer-lived connections, using media links to facilitate interaction between workers in multi-site research institutions. The social interaction system took place at the Xerox PARC Media Spaces [15] [16] connected researchers at Palo Alto and Portland through mobile video and audio links. The media ran over two years, seven days, 24 hours until the office closed. It was initially intended for supporting formal meetings, though, the majority of interaction were chance encounters lasting less than five minutes. A similar work at Bellcore Labs, the VideoWindow [17], connected researchers in two different floors of a building using media tools and large projected displays in common areas. Also, Microsoft Virtual Kitchen [18] linked three kitchens using media connection that showed each kitchen a view of the other two kitchens, as well as the local view; in space, they were showing television channels to attract viewer attention. This experiment additionally provided privacy protection for the users; the media were turned off temporarily at the user's will [13]. Similarly, Telemurals [19] used a more abstract media connection initiating interaction between two remote spaces. It was trialled with the focus of 'events social catalysts'—they performed image processing technique to overlap two video feeds coming from different physical locations into one image. The term stands for the phenomenon which people become more likely engaging to the installation by seeing notable events (i.e. graphic effects) implemented into the design of the communication medium.

During the 1980s and 1990s, there were more ubiquitous systems targeting at indoor environmental settings as opposed to installed in outdoor settings with the concepts of public participatory and urban installations. In addition to projects introduced earlier, the enormous amount of research in 'pervasive displays' has attempted to improve communication efficiency between coworkers and researchers who are at the physically remote locations. Instead, in the early to middle 2000s, projects began to emerge which the domain focus expands to the public interactions using the display as a medium.

The Interactive Wall Map [20] adopted a world map with public displays embedded to promote interaction between passersby; world map naturally attracted people passing by and sparked communication between them. The wall was covered by a large map (approximately 4m x 2.5m) augmented with three pairs of flat-panel touch screen monitors within different geographic regions, and 24 buttons were placed over cities. Pressing the button showed information based on the geo-location (city), and the system could also sense the presence of user via infrared badge system (seen in many practices used to automate computer response towards human presence in the interior of building back in this time) [13].

The Opinionizer [21] was designed for informal gatherings (social setting) which aim for the collective building up of shared contents that are input by people nearby and then showed to everyone in the room to elicit communication between strangers. The Opinionizer allowed anyone to type in opinions and comments, which then appeared in the shared display so that other people may initiate a conversation with the content shown. The system designers intended building a system that users feel comfortable and enticed to take part in rather than be wary of. The system was trialled in two scenarios; a book launch party (two hours, approximately 300 attendees) and a welcome party for new postgraduate students within one school of a university (2.5-hour deployment, approximately 150 attendees) [13].

More recent research on long-lived public display deployment and urban computing have emerged from a variety of locations, including city centres, rural communities, and university campus [13]; which they typically used it for advertisement, social interaction, community participation, and research purposes.

In the late 2000s, A network of interactive public displays called "Ubi-Hotspots" [22] [23] was deployed in Oulu, Finland. The network consists of twelve hotspot sites, each including one or two 57-inch LCD panels (indoor hotspots feature a single outward-facing display, while outdoor hotspots use two back-to-back LCDs to support use from both sides). Each hotspot is also equipped with a loudspeaker, cameras, network access points and an NFC reader [13]. Ojala et al. promoted the media as 'heavyweight' urban probes which enable long-run research activities and ultimately “making the urban space a better place for people” [22]. The hotspots were used as application platforms accommodating web-based applications developed in compliance with minimum guidelines for the design layout; a few web applications were initially deployed and examined along with the usage of the hardware installation. Years have passed now, the system is still in operation mode, being trialled with new applications and evaluated the status in the multi-level perspective [23]. Currently, the installation is in the phase of turning into a community fabric—community members take the technology intervened in the city for granted.

2.1.2. Thematic categorization of prior researches in urban computing

2.1.2.1. Community memory and community consultation

The DIY Shrine’ built by a company called Umbershoot based at Melbourne, Australia, was placed at Federation Square in Melbourne during their 2014 annual winter festival; the deployment lasted for 22-days during the festival, with the theme of ‘shrines’. The installation used iPad and GoPro to display questions and record video of the user; questions were designed in such a way that prompts the users to share their confessions inside the phone-booth-like structure. Space inside the structure sized for two adults to fit in, and designed so that participants feel comfortable sharing meaningful and personal stories with the DIY Shrine. In the end, they shared the video record on Vimeo channel and in the urban screen; the content was rich and authentic as attributed to the intimate space setting that provoked people to share genuine stories [8].

The ‘City and You: Tell your story and reconstruct the city’ was a project carried out with the aim of building a cityscape using text and video of memories from people in the city. The installation took place around the city of Ahmedabad, India, and consisted of a rotary telephone on a desk, wired with a computer. When the phone was picked up by a user, a voice asks questions about the city and its meanings to them; the city formed the audio content into stories of text and video to share with the world through social media [8] [24].

2.1.2.2. Urban intervention and urban computing

The ‘N Building’ [25] was designed by Terada Design ARCHITECTS located at the shopping district of Tachikawa station in Japan. The intent of the project was to create an urban intervention (building) which collects quality information from community members while keeping them unhindered by the identity of installation or ubiquitous signage. The façade of building displayed a QR code allowing interactions with users outside in two different levels; the first level interaction provided a link to the latest shopping information, while the second added an augmented reality component to the façade that allowed users to view the tweets posted by other users. The media facilitated communication between users inside and outside of the building by using tweets collected from their website [8].

The ‘CITYtalking’ [24] was developed by an action research/performer Astra Howard for her PhD ‘Orchestrating the public: to reveal and activate through design the experience of the city’. It is a movable booth construction that wheeled around Melbourne streets for five weeks during 2006. The booth featured two compartments, one for the participant and another for herself. An LED screen attached to the façade of the booth displayed an instant translation of narratives provided by the participant. An intercom between two compartments allowed participants to share stories of theirs anonymously; Howard then quickly published the stories using outside LED screen for passersby to read. Her another project called ‘City Feedback’ took place in the city of New York in 2008. 30 yellow cards were placed in thirty public telephone booths throughout the city. Each was printed heading of ‘New York’. Her intention was to let passersby pick up those cards and use them for describing their experience of the city before dropping it into a mailbox; the destination of address and recipient were filled in advance to return to Howard [8].

2.1.2.3. Combination of multiple concepts

The ‘InstaBooth’ [8] [2] is a community engagement platform with multiple urban computing design concepts combined: placemaking, participatory design, and community memory. The report highlighted the value of multi-disciplinary design team consisting of academics and practitioners with backgrounds in architecture, interaction and visual design, HCI, urban informatics, lighting designers, media artists, and much more. The InstaBooth provided communication modules to facilitate the creation of dialogue and sharing ideas; these modules are various that there are ways of contribution to the system from writing a note to giving a hug. The system has been deployed around the city of Brisbane, Southeast Queensland, Australia for seven times already, though, it is still in the phase of prototyping combining features. Nonetheless, some of the valuable lessons learnt from their early deployment are as follows:

1. The system should be used in various ways and by users with diverse attitudes (in all different individual ways).
2. The system would become more meaningful when users gradually take the ownership of the installation: it grows not only by the designers and researchers but also by allowing users find value in it, thereby, these ‘agents’ will become the part of the community [2] [26].

2.2. Virtual Worlds & Virtual Environments

The Virtual Environment (VE) and Virtual World (VW) have been traced back to their first emergence much in the past than the corresponding Virtual Reality (VR) technologies (also referred as ‘metaverse’ coined by Stephenson (1992) [27] [28] [29]). The first concept of VW was conceived by Carroll’s book *Through the looking-glass* (1883) [30]; since then, possibilities and potentialities derived from the novel technology have hitherto been growing.

The VE, one of the core elements in VR or VW, focuses substantially on the visual perception of the users than any other sensory or non-sensory requirements appearing in a VR system. What two differences between VR or VE to VW then? From the visual point of view, superficially, these technologies seem analogue to one another, because they all present a computationally synthesised space; particularly the boundary between VE and VW is found more ambiguous than that of VS. The potential confusion appears for certain circumstances which a VW has no indigenous characteristics than just a persistent mediated environment. The rationale for why we should clarify the definition can be described twofold: first, to set the social implications of technologies so that we could contribute to the society and research by using appropriate terms; and second, to better navigate conducting researches [31]. With this respect, Schroeder (2008) has an example of the definition of word ‘virtual’: “it has come to mean anything online (as in ‘virtual money’) which the word becomes useless from the point of research view as it is too broad” [31]. Then, he proposes his idea of differentiation for VE or Massively Multiplayer Online Roleplaying Games (MMORPGs) from VW: VW is meant to be a place for socializing, a social space in which people mingle and converse with each other; users should be able to experience a VE on-going over time; the VE has to accommodate a large population. The online games (also referred as ‘structured 3D VW’ [6]) and VE as opposed to VW, are the subset of VW apart from socialising and gaming aspects (see Figure 1). Although, online games are also used for socialising in some cases [32], primarily they are designed for gaming purposes that involve gaining points and competing with other players [31]. Figure 1 shows an insight towards these terminologies taxonomically from Schroeder’s viewpoint, which regards their properties and associated technologies. VW, VR, and online games converge on a single shared point in common that is VE, while the rest parts of each technology retain their uniqueness of the class. The taxonomy figure below comes across the best with examples: a VW which can operate without any properties of online game or/and hardware devices of VR; an online game which can work without an element of socialising. Similarly, VR, VE, online game and VW all have partially overlapping fields; for instance, a VR system for gaming purpose, a VW complementarily equipped with devices used in VR systems—commonly a headphone is used for communication purpose, yet it is still social-oriented VW application.

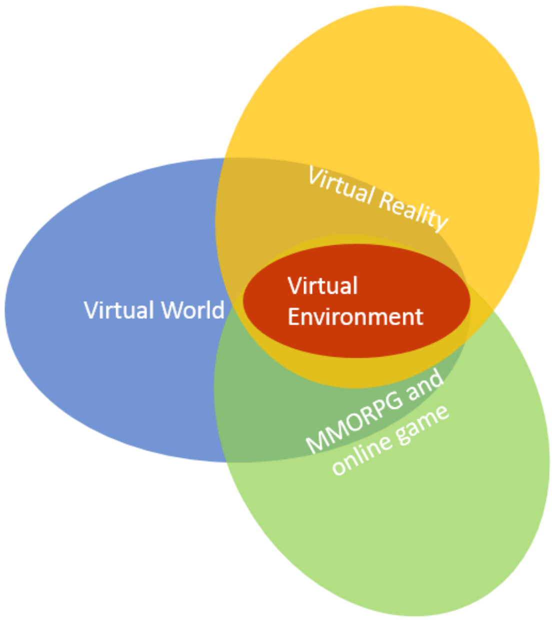


Figure 1. Definition of VR, VW, VE, and online games

2.2.1. Applications and contents

Since the computer era has begun, software and hardware technologies have developed alongside. Computers rapidly invaded into our daily environments, changed and advanced the quality of human life. By the result, the majority of our activity domains nowadays is vitally engaged with computer applications running on the machines around us. A contemporary human life thereby has become which hardly can sustain alone without the help of computers and their applications. Moreover, such machines being evolved smaller, lighter, and even more powerful than ever before, have led us to the era of ubiquitous environments in which the systems are impalpable by human's natural senses.

The prospective target groups for VR applications are identified as a personal computer, mobile device, and console game. These application platforms are already eagerly anticipated by their users from all over the world, though, for them to be successful, they need to work together with compelling contents. Some of the practices trialled in the literature are presented by Gobbetti and Scateni (1998): virtual prototyping, simulators, training, and augmented realities. Gobbetti and Scateni stress that the VR applications deployed in the field of architecture have been one of the most successful cases over the history [33]. These applications enabled remote collaboration in the project work which has led to allowing architects build prototypes in the very close vicinity to the clients based on their desiderata [34] [35] [36]. Similarly, for the time this thesis is being written, July 2016, a mobile application called "Pokemon Go" has been introduced by a company called "Niantic". The game is by far epoch-making regarding the number of users registered in a short period and the concurrent play users. This 'craze' was catalysed by combining the concept of AR technology that is relatively fresh to the publics with the globally favoured animation characters which they perfectly blended into a mobile device. The phenomenon was remarkable as to the implications of which an ordinary mobile application can become this famous in short. Taking Pokemon Go as a reference, we argue that creating a novel application by involving a VR or other family technologies can turn out far exciting than we could ever imagine.

2.2.1.1. 3D city modeling in VEs

Digitalization of analogue city data (also called open data) into 2D has started since the 1980s, and subsequently, 3D from 1997 when digital maps and computers were popularized [37]. Throughout the conversion process, each transition required a reinforcement of the existing data quality or/and a collection of new data. These can be expensive operations dealing with large-scale city models, as opposed to small-scale or temporal modelling project; the modelling is done by historical research on data preserved in the form of pictures or maps [38]. A means of collecting data for geometry and texture construction of the city with large-scale consists of laser scanning and photogrammetry [39], but these approaches still require a vast amount of manual work using 3D modelling software tools [40]. In the recent past, to tackle these problems, collaborative modelling has emerged. This involves a large number of participants or a particular crowd of people to aid existing methodologies. For example, Palmquist and Shaw (2008) [41] trialled the collaboration modelling method with students, who were given a high-level task guideline. Maher et al. [42] studied the collaborative phenomenon and motivation for crowdsourcing in the design domain. Furthermore, Virtanen et al. [43] and Alatalo et al. [40] demonstrated it at the platform level, in which the virtual city scene was accessible by a larger number of publics and researchers through Meshmoon platform, and Uden and Zipf (2013) [44] also practiced developing an open city modeling platform in which the 3D modeling task is partially contributed by crowdsourcing.

The use case of 3D city model is multifold, and depending on which purpose, requirements are also varying. Carrozzino et al. (2009) enumerate examples in the application domains: urban planning, communication, emergency management, architecture, documentation, social networking, education, etc. [38]. Along the lines, Döllner et al. stated that the applications of city model could benefit IT solutions, such as telecommunication, disaster management, homeland security, facility management, real estate portals, logistics, as well as for entertainment and educational purposes [39]. And the state of the art review has identified at least 29 use cases, more than 100 applications from the multidisciplinary fields [45].

Regarding requirements for 3D city models, Carrozzino et al. [38] state that urban planning requires building models equipped with datasets of high precision geo-reference and spatial-volumetric dimension that correspond to the real-world counterparts. Typically, architectural applications generate intensive requirements in visual details and realistic models, whereas social applications demand 3D models to be immersive and appearing by focusing on aesthetic quality of the environment, so that attracts more users [46].

The history of Meshmoon has begun with an open source software project titled as RealXtend, which produced a software called Tundra and Naali viewer. These technologies have goals of extensible, widespread, and freely accessible virtual application platform. Afterwards, the technology has extended the implementation into underlying core technology of the Meshmoon. Tundra is an entity-component model platform for VR applications which facilitates developers in creating VEs for any purposes. The technology supports essential features for building VE components (i.e. avatar). It is under Apache License v2 and was commercially published as Meshmoon. The Naali name came from a Finnish word meaning arctic fox, indicates the Finland origin of the technology, which pursues the idea of widespread availability of the Firefox Web browser. It is a client viewer for virtual environments that can connect to SL, OpenSimulator—open source multi-platform, multi-user 3D virtual world server to which virtual clients such as Naali, or SL client can connect—and Tundra server. Also, Naali can run on Windows, Linux, Mac, and some mobile platforms too. Tundra resembles SL and OpenSimulator, that are at the same time, competitor software technologies from which Tundra was inspired [47].

2.2.2. Second Life and its applications

There have been many web-based VW platforms released to date, which either pursues a full-reflection of the Real World (RW) with no physical limitation or looks at the entertainment aspect of a gaming platform. Examples include There.com, Kaneva, Active World, Minecraft, Second Life (SL), and World of Warcraft (WOW) [28] [6] [27].

Despite the overwhelming popularity of software titles such as Minecraft, which holds a larger number of users attributed by the easy-to-play factor, SL is still arguably the most popular web-based virtual world platform with the largest population of more than 10,000,000 residents [28] [48].

Second Life (also labeled as MUVEs—multi-user virtual environments [6]) project was inspired by Stephenson's book *Snow Crash* (1992) [29] [49]. The software holds integrated properties of entertainment, social networking, business market, media, and advertisement across the applications which are previously or/and currently in service [28]. In particular, education/training [49] [50] [51] [52] [53] [54], healthcare/medical, and simulation [55] are broadly applied sectors at enterprises, institutions, and organizations.

2.2.2.1. SL as a medium for education environment

More than 70 colleges and universities have built campuses or classrooms recently in SL [27] [54]. These VWs share a dedicated goal on creating an online learning environment referred to as Massively Multi-learner Online Learning Environment (MMOLE), or Virtual Learning Worlds (VLWs) [27]. The institutions that have built a learning environment in SL seek to forge a learn-anywhere environment through VE space; they believe the demands for establishing an unconventional educational culture are crucial in response to the emergence of 'digital native'—the term refers to a young generation who has grown up surrounded by advanced digital technologies as opposed to the 'digital immigrant' who was born before the digital age [56] [49]. Today's VR systems are rather uncertain, complex, and noneconomic for the education purposes; however, VWs such as SL promise a compelling future methodology for education by offering everyone opportunity to participate in a low-cost educational setup [49]. In the pedagogy point of view, VE and VR technologies could provide a prominent advantage in supporting educational activities through the realisation of constructivist learning curriculum—a.k.a 'experimental learning' by Jarmon [52], meaning that learning-by-doing, as students are expected to learn by assimilation [50].

Many researchers have indicated with experiments that the educational activities conducted in VW can drive significant consequences that are promising in terms of evolutional environment for teaching and learning activities [49] [51] [52] [53]. Though, still, some argue that the VWs without VR components will suffer from poor user experience on account of existing limitations. In fact, the cases observed from several works of literature showed that as the relatively new technology is used for learning and teaching activities, some participants had to undergo a process of learning the technology to fulfil the tasks during the experiment [49] [53] [52].

Witmer & Singer [5] provide an insight into the future computer metaphor in VEs employed for learning and teaching: if the current desktop metaphor is perceived to human as 'from the outside in', the perception of future metaphor should shift to 'from the inside out'. This way, the learning ability will improve as the user becomes an integral part of the stimulus flow, and thus, the immersive environments may become better educational tools than standard computer-based learning systems; Here the phrases 'from the outside in' and 'from the inside out' mean respectively "how the current perception of using computers to us" and "how the future perception of that should turn into", implying that users should be able to perceive themselves inside the VE rather than outside sitting at terminal to communicate with computer in the future.

2.2.2.2. SL as a testbed for mock simulations

As mentioned earlier, applications for simulation purposes have also shown a high influence on various industry sectors such as business/marketing [28] [57], academic [58], medical [54], Especially, where the physical limitations causing difficulties in building and controlling of primary vehicle, ship, and aircraft [59] industries have led one of the first in-practice employment of VW on the simulation application [60] [55] (see Carr & England, 1996, for more details for the prior practices of VW [55]).

Veksler [61] reported that as a simulation environment for cognitive modelling, SL showed the best performance—performance in environmental fidelity on a proper balance between simple simulation environment and complexity and constraints of RW—over many other virtual simulation environments including RW simulations.

Similarly, as the technology favours enterprises with the provision of an affordable testbed for their new promotions and distribution of products, SL has become an active business and marketing platform for many big brands such as IBM, Coke, Adidas, Sun Microsystems, Toyota, Sony, and more [62] [63]. These companies seeking to take more shares than other competitors in the online market industry have found a perfect experimental environment in SL, in which its users might be already online consumers [28].

2.2.3. Telepresence and Teleoperation

In the nascent exploration in VR technology, there was no effort made for defining which system can be called VR as opposed to the other media types without delineating the hardware components employed in the system [64]. In the still early 1990s, the term 'Telepresent' coined by Marvin Minsky [65]. The term was first introduced to VR in reference to teleoperation systems for the remote manipulation of physical objects. Subsequently, Sheridan [66] [67] adopted 'Presence' to describe the human's perception of being in an environment that is physically remote from the user, or mediated by a computer, or artificial; in the meanwhile, he reserved telepresence only for cases involving teleoperation. On the other hand, Held and Durlach [68] use telepresence for the common experience perceived from VEs and teleoperation in the same section of the journal [64].

Steuer [64] argues that if the traditional process of communication was defined by information transmission between sender and receiver who are independent individuals, telepresence view focuses on the relationship of an individual who is both sender and receiver and the mediated environment which the person presents in [66]. He also defines determinants of telepresence as twofold: vividness and interactivity (see Figure 2). Vividness is an ability to which the technology is capable of creating a sensorily rich mediated environment. Similarly, Interactivity is defined by the extent to which users can influence objects or situations in the mediated environment. However, Steuer stated that the definition of interactivity differs widely between researchers; thus, it is still an open subject [64] [46].

Figure 2 draws a hierarchical relation tree indicating the subsets of variables that constitute telepresence. These determinative properties of the stimulus are, however, not identical across the range of perceivers as the VRs reside in their consciousness; for instance, an absence of auditory information in a mediated environment might be a crucial hindrance for someone, whereas for others it might be even hardly noticeable [8] [46].

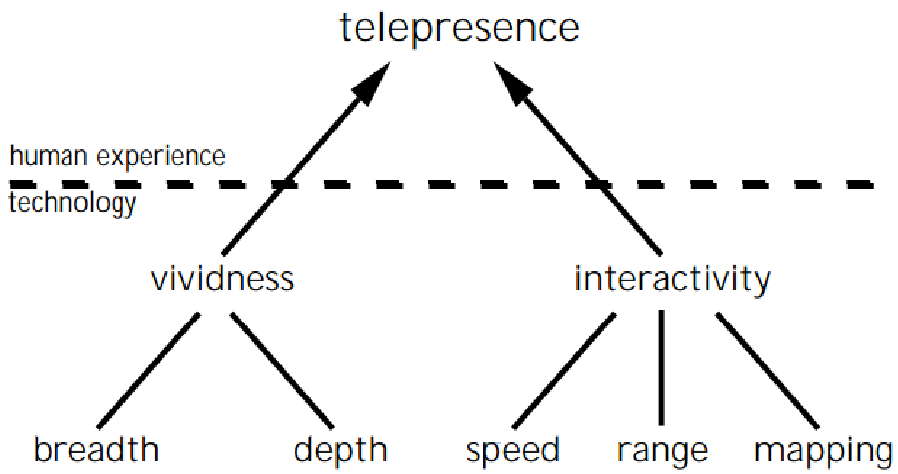


Figure 2. Technological variables influencing telepresence from Steuer (1992)

Sheridan [66] identifies five variables of a medium that influence the sense of telepresence from the technological and context perspectives.

The generalised two determinants that contribute to enhancing vividness are 'breadth' and 'depth'. The breadth defined by J.J Gibson [64] refers to stimulus-driven sensory components that consist of distinct perceptual systems. The identified perceptual systems are fivefold: auditory system, haptic (touch) system, taste-smell system, and visual system [69]. Most of the traditional media possess the above attribute though, they are unlikely to meet the variety of systems, which results in a weak vividness of the application. Each input system (sensory modality) can enhance the vividness of the technology, and its redundancy—combination of sensory systems, also referred as 'Extensive' [46]—can substantially improve telepresence. Similarly, the depth, as opposed to the breadth that enhances vividness by adding more sensory attributes, concerns with the quality of the media. In other words, the technological displays of the meditated environment and aspects of hardware grade contribute to the depth, such as pixel resolution, graphical richness, quality of the environment [64].

'Interactivity' is another important technological variable which consists of speed, range, and mapping of the technology. Interactivity contributes to inducing the sense of telepresence with the help of developing the subset variables of the system. The definitions alter from system to system, or situation to situation, though. We describe each determinant as follows (more details are found from [64] [70] [71]): speed is determined by immediacy, that is an ability of technology for creating an environment that implements immediate changes occurred in RW, or vice versa (if applicable). Range is defined by the amount of mediated environmental volume that can be modified/adjusted by the RW inputs. Lastly the mapping—or 'Matching', an alternative term referenced by Slater & Wilbur—refers to availability which movements of the user are mapped into the actions occurring in VE, or are matched with proprioceptive feedback [71] [70] [46]. The mapping can be very arbitrary; by way of example, a human gesture that has no relation to which actions are occurring in VE.

2.2.4. Immersion and Presence

Definitions for ‘Presence’ and 'Immersion' are logically distinguishable based on the domain knowledge of human science and psychology. immersion is addressed as an objective constituent to the human psychology, whereas the presence—a.k.a tele-presence—is a subjective experience of an individual [72] [60] [5] [46]. Presence is a subjective sensation or mental manifestation that is not readily amenable to the objective physiological definition and measurement; the more the sensation is strong, the more the experience is meaningful [66]. In contrast, immersion can be objectively assessed other than how humans perceive it; an example is given by Slater [73]: if the immersion is analogous to wavelength distribution that is the objective measure for describing the colour of an object, presence is analogous to the perception of the colour from the human (thinking of metamer, a colour may be perceived differently depending on the individual, or illumination).

Given distinctions of two psychological experiences, it is claimed that immersion can be concerned as one of the key factors to the high degree of presence [60]. The statement is supported by Witmer & Singer [5] who developed the questionnaires for the measurement of immersion and presence quantitatively, say “individuals who have a greater tendency to become involved in a variety of activities as measured by the ITQ, should report more presence on the PQ” (ITC: Immersion Tendency Questionnaire, PQ: Presence Questionnaire) (ITC: Immersion Tendency Questionnaire, PQ: Presence Questionnaire).

A set of attributable properties of presence was investigated hierarchically in Figure 2, for which each property presents with a brief definition of it. Additionally, Slater & Wilbur [46] complement the list by introducing another factor variable called ‘Plot’. It is the extent to which the VE in a particular context presents a self-contained storyline which has own dynamic and displays events that are distinct from anything around the real-world context; this can potentially remove the user's awareness to the real-world and promote the concentration in the mediated environment with its ecosystem. The notion includes Zeltzer’s (1992) ‘Autonomy’ and 'Interaction' for which each represents objects have unique behaviour and the extent to which the degree how the user can contribute to unfolding events taking place in VE respectively [74].

People are more excited and focused on the activity and the environment in VE when they face a novelness in it; the psychological status of a human with a thorough focus in their motions which requires novelty, immediacy, and uniqueness of the experience enables the individual to be broadly aware of the task environment [5]. Similarly, Fontaine’s [75] studies in general presence linked to VEs stated, the broad focus is a necessary for achieving a high sense of presence in VEs. Also, the attribute of the meaningfulness of information showed more attention was guided toward the experiment, which attributes to ‘Selective Attention’—willingness of person to ignore external distractions while focusing on something [76] [77].

Lastly, according to Witmer & Singer [5] [73] and the contents presented in ITC, sense of immersion is derived from one’s intrinsic characteristics toward a tendency for immersing him/herself into certain activities that are currently performed or the environment in which he/she are living. In the context of VRs, however, the immersion stands for simply what the technology delivers from an objective point of view; namely, if the system possesses all sensory modalities and tracking that preserves the fidelity about RW counterparts, it is assumed to draw a more immersive result in the measurement.

2.2.4.1. Immersion and presence in relation to task performance

From a training point of view, the efficacy of RW tasks performed in a VE is substantially dependent on the presence of the user [75] [66].

Witmer & Singer [5] suggested that many factors which enhance the presence are also instrumental in supporting education activities and task performance in VE; Similarily, Lucia, A. et al. [49]stated that the degree of presence has a strong correlation with learning performance; However, despite all these; it is still rather rash to affirm that the task performance is directly affected by the presence and immersion because there have not been enough systematic studies conducted to underpin the argument logically. A few studies in the context have carried out experiments which reported findings that statistically significant correlations between PQ scores and performance on the simple psychomotor tasks performed in a VE were observed [7] [78]. However, they failed in a subsequent experiment, which conducted under a different setting while keeping the task unchanged [79], therefore, the results are unreliable. Although a streamlined interaction, pictorial realism, social interaction, and optimised system factors may increase the degree of presence in a VE, that cannot assure the improved performance of the user. And so, it is rather ill-advised to guide one to achieve the high presence and immersion as a means of obtaining a better performance on a given tasks [72].

3. VIRTUAL SOAPBOX SYSTEM

Virtual Soapbox System (VSS) is a web-based VE enabling the Soapbox System (SS) to build a hybrid community platform in which users who are physically remote from each other are 'being connected', or feel 'sharing the space'. VSS as part of SS accommodates online users whose participation to the SS is not dependent on geo-location. While the primary intent of VSS is to reach people online, offering them a means of contribution to the community discussion, the system also focuses on bridging diverse individuals into the community. It is noted by Jenkins [4] that the more vocal citizens are not necessarily the most representative of the community, which can be taken as an insight in building community participatory architectures that the involved system designers need to take more heed of the less-vocal population.

3.1. Soapbox System

The soapbox system employs VE technology as a medium for a shared community platform which allows users to present their thoughts and ideas under the design concepts of community participation, community consultation, and urban intervention. The SS breaks down into three sub-systems: two applications and one middleware server, in which user participation to SS takes place in two different ways: using Physical Soapbox System (PSS) on site situated at a public place, or taking advantage of web technology with VSS. The SS aims at encouraging people to take part in the social activities through on-/offline soapbox systems; communication between two application systems is enabled by the Middleware that navigates media links coming from PSS and VSS. The objective of SS lies in building a 'common ground', for which the community members can join to share, discuss, exchange, and converse opinions and ideas. Eventually, we hope to hand over the ownership of the system to the users such that the system grows onward as a fabric of the community.

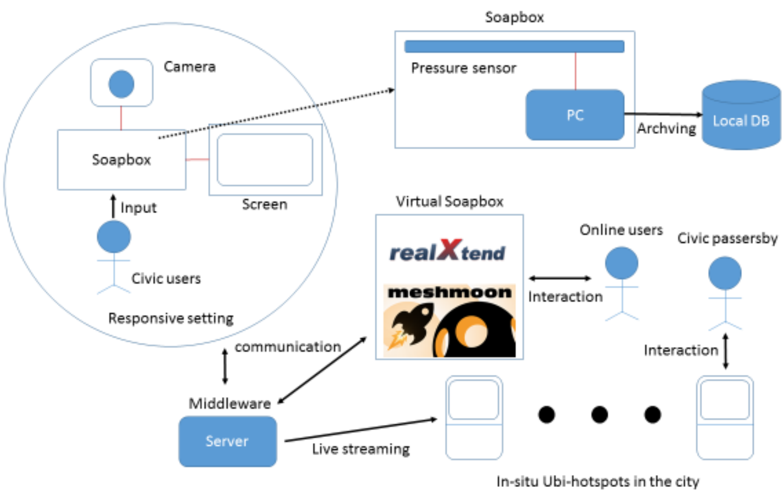


Figure 3. High level architecture of Soapbox System: the design describes systematic collaboration of systems across PSS, Hotstpots, Middleware, and VSS.

3.2. Virtual Soapbox System Design

The VSS project was initiated in June 2015 in conjunction with another two sub-systems as part of the SS. Throughout the system design phase, we devoted an appropriate amount of time to design the system structure and plan for the future implementation. With given fundamental concepts of the system, suggested by Hannu Kukka who is one of the project providers as well as the supervisor for this thesis, we first divided SS into three sub-systems and assigned to each of the team members.

Figure 4. Primary requirements for Soapbox System, and the discussion to workload division

3.2.1. Research on system design goals

The research phase was dedicated towards a breadth wise investigation of the requirements domain. During this phase, we focused on the following:

1. Learn stakeholders, who are they?
    a. Identify primary and secondary stakeholder
2. Discover Goals
    a. What are the requirements?
    b. What are the research goals?
3. What are the technologies we need?
    a. Do current technologies support the system?
4. What are the expected outcomes?
5. How is it done in other related software?

Some of the above questions were considered as within-team deliverables for discussion at the team meetings. All of the members participated and contributed equally to above components that affect the structure of SS, though, individual system parts were scrutinised, filtered, refined, and finally concluded respectively.

In the end, all system designs were put together in before the implementation phase, being contemplated in a sense that the whole system generates a cohesive result for the SS.

3.2.2. Strategy

If the previous phase had focused on the design of sub-system individually and the features that are relevant to the system, in this phase, we were to ideate. Not all the ideas brought to the meeting table were pertinent to the SS; therefore, we brainstormed to isolate components in accordance to relevance, discussed systems integration, and measured how to cooperate in building features were required by two different system components (PSS and VSS). Also, it was during this phase we conducted a literature review and built UML diagrams.

3.2.3. Requirements analysis

During the system analysis, we identified the stakeholders of our system, statement of requirements, and 3D model designs in an exploratory manner; the stakeholders were selected based on our research agenda. Through the following chapters, we will present these contents in detail.

3.2.3.1. Stakeholders/Users

1. Direct users: people who want to deliver a speech using the system.
    a. Student representatives
    b. Institutional/organizational representatives
    c. Educators
2. Indirect users: people who want to propose or share thoughts and ideas to others (Civic Engagement)
    a. Ordinary community members (on-/offline)

The research goals for this thesis place emphasis on the performance in delivering a speech and therefore, the target group of our interest becomes those who have more chance to speak to the public.

Besides, as the original purpose of Soapbox aimed for the general civic users, everyone in the community also can fall into a potential user group (indirect stakeholders).

3.2.3.2. Statement of requirements (SORs)

As noted by supervisors during the discussion session for the workload division (Figure 4), the primary requirement for VSS boils down to the following: build a VW that facilitates online users join a community discussion, and deliver public speeches. That being said, we have formed a statement of requirements for VSS based on UML diagrams and research objective. Although the list has changed over the development process, we present below of the outline upon which we built a guideline for the system implementation. The term 'real-audience' here refers to a public audience engaged with the system.

1. Build a virtual space of multi-user-participatory platform that resembles the City of Oulu.
2. Design a 'responsive' virtual space equipped with 3D user-interaction models.
3. Establish media links with Middleware system to exchange video feeds both outbound and inbound.
4. Enable features for the speech activities (begin, vote, and comment a speech).

The list above may be generic and contextually incomprehensible; hence, we boiler-plated them in a user-centered manner. Of its outcome will delineate the relationship between system functionalities and stakeholders, which will aid understanding of the functional components of each system.

The boilerplate used is <*User*> shall be able to <*Capability*><*Performance or constrains*>. <*Capabilities*> are the functional requirements of the system, while <*Performance or Constraints*> are non-functional or performance requirements of the system. With these in mind, the boiler-plated statement of requirements specifying the functional and non-functional requirements are as follows.

1. The <*online users*> shall be able to <*functional requirement*> < *non-functional or performance requirement*>
    a. The <*online users*> shall be able to <*begin a speech*> <*if the state of speech is inactive*>
    b. The <*online users*> shall be able to <*vote on the speech*> <*if the state of speech is active*>
    c. The <*online users*> shall be able to <*report on the speech*> <*if the state of speech is active*>
    d. The <*online users*> shall be able to <*comment on the speech*> <*if the state of speech is active*>
    e. The <*online users*> shall be able to <*view real-time state of the vote and comments*> <*if the state of speech is active*>
    f. The <*online users*> shall be able to <*register a speech*> <*if the state of speech is inactive*>
    g. The <*online users*> shall be able to <*view the speaker*> <*if the state of speech is active*>
    h. The <*online users*> shall be able to <*see 3D screens varying in size*> <*when the number of active users in the venue is above/below the threshold*>
    i. The <*online users*> shall be able to <*view the real audience in real-time*> <*if the speech and PSS is active*>
    j. The <*online users*> shall be able to <*hear to the speech*> <*if the state of speech is active*>
    k. The <*online users*> shall be able to <*be informed about a new speech begin*> <*if a speaker begins speech*>
    l. The <*online users*> shall be able to <*view the remaining time of ongoing speech*> <*if the state of speech is active*>

   m. The <*online users*> shall be able to <*teleport to the venue*> <*if speech is active*>
2. The <*online speakers*> shall be able to <*functional requirement*> <*non-functional or performance requirement*>
   a. The <*online speakers*> shall be able to <*view the state of the vote in real-time*> <*always*>
   b. The <*online speakers*> shall be able to <*end the speech*> <*if they wish to*>
   c. The <*online speakers*> shall be able to <*view remaining time of the speech*> <*always*>

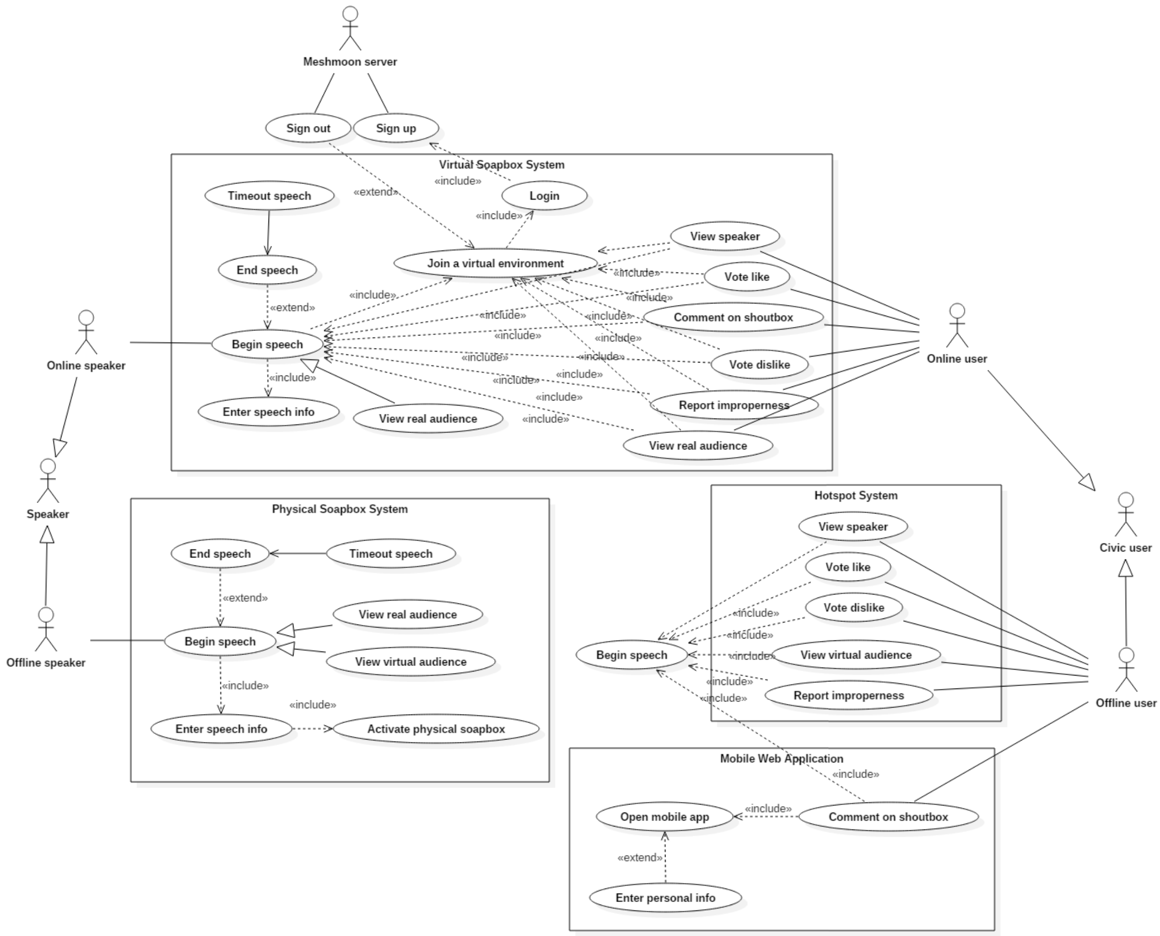


Figure 5. Soapbox System: User-centered use case diagram

The statement of requirements and use case diagram illustrate system responsibilities associating different stakeholders. Subsequently, non-/functional system activities in SORs are further analysed by the activity diagrams presented in Figure 6, 7, and 8. Likewise, what a flow chart does, activity diagram draws a flow of actions in a sequential order; it comprises of some of the main features of VSS. Figure 6 and 7 illustrate a speech begin and end respectively; Figure 8 describes the voting process.

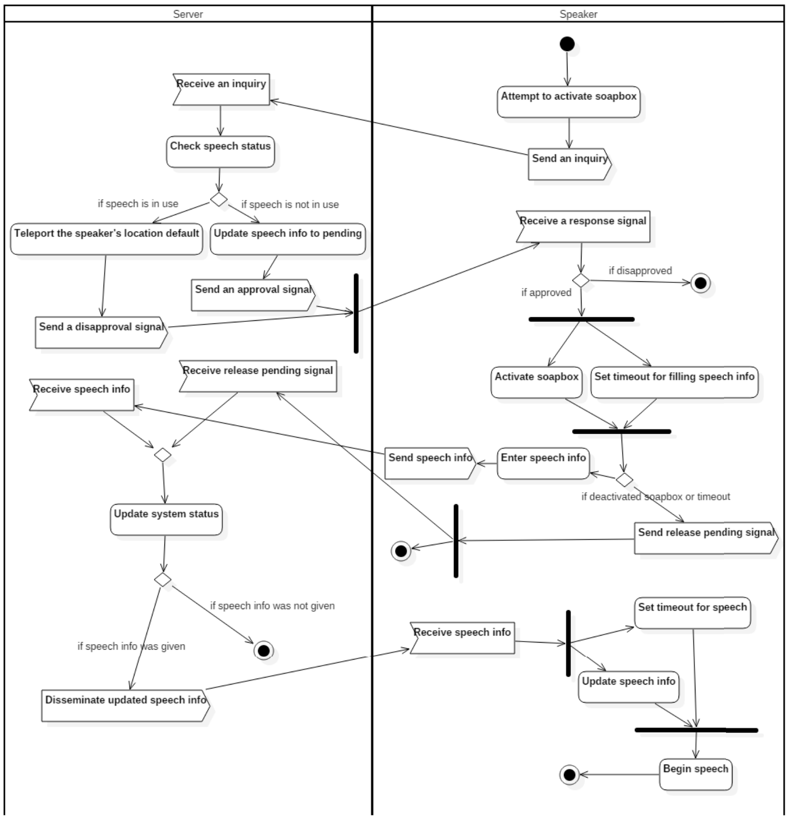


Figure 6. Activity diagram shows action flow within the system when commencing a new speech

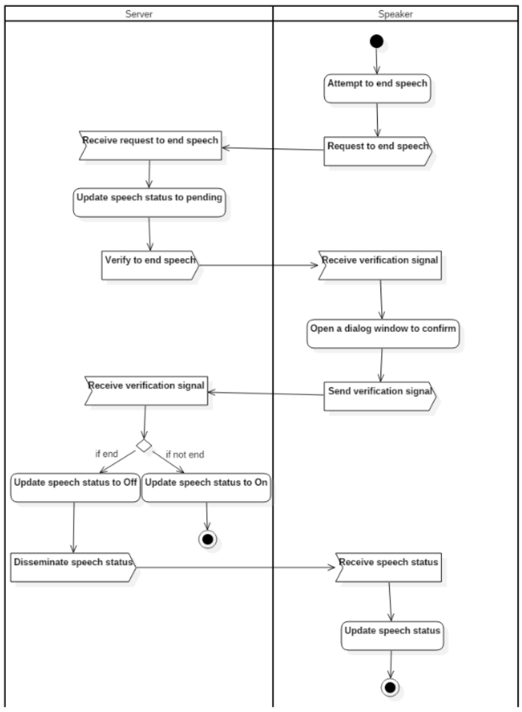


Figure 7. Activity diagrams: a speaker attempts to end the speech

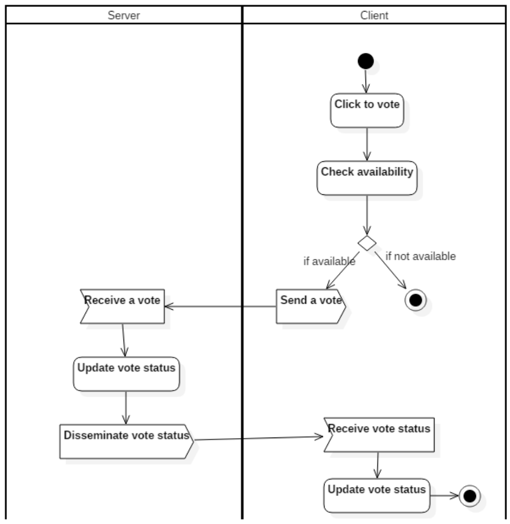


Figure 8. Activity diagrams: an online user intent to vote on a speech.

3.2.3.3. Environmental design requirements

The virtual environment was designed to be an accurate representation of the corresponding physical space. As noted earlier, natural interaction and control, redundancy of sensory modalities in the system, aesthetic visual quality, and perception of self-inclusion—indicates the extent to which physical reality is shut out for a user—are the critical factors for enhancing the sense of immersion and presence in VE [5] [38] [46]. Keeping that in mind, we have designed a textured wooden soapbox that resembles the physical one (see Figure 9), and the way a user interacts with it makes the whole interaction natural. The soapbox reacts to a tentative speaker stepping on the box. Also, the future evaluation has attributed to the design of VE; we plan the deployment of SS to take place at downtown Oulu. That being said, the surroundings such as buildings and artefacts in the vicinity of the venue are crafted as analogue as possible to the real-life designs—for the evaluation presented in this thesis, however, we use a new model of university campus for the study. We designed 3D objects uniformly between PSS and VSS that we expect from such inherent coherency around the environment remove difference across the SS systems.

Instead of a more immersive object design for an interaction tool, we followed the conventional means of displaying information, using an add-on dashboard that has been adopted in many web pages. The rationale behind this lies in the statement that achieving a high degree of immersion and presence is not the crucial factor for our research interest (i.e. we do not use VR hardware devices) [72] (see '<u>Research questions and methods</u>'). The assertion is also supported by a study conducted by Spagnolli and Gamberini [80] saying "Would the visibility of technical medium really cause 'emersion' of the user?"—'emersion' here stands for a psychological condition of the user which the immersed sensation into the VE is snapped out by an occurrence of an in-/external disturbance or intervene. They conclude that, the technical dysfunction and perceptibility of the system medium as such are not determinative factors for emersion. Also, the attempt to display all the information into the VE without a trace of the technical medium is usually a non-trivial task. Lucia et al. [49] showed it was viable as they had the information about the number of students are participating, and it was relatively small. Given the nature of dynamics, openness, and accessibility of an open VE, we chose the traditional dashboard design.

Figure 9. Soapbox design, left: created with Blender for this project, right: real soapbox image [81].

3.2.4. System design

We are building a web-based VE that is capable of streaming media properties of online users real-time. The system will enable users—our primary and secondary stakeholders—(a) to practice public speeches, (b) to share ideas and thoughts with others, (c) to listen to other speeches. It is a challenging project concerning its novelty, which connects RW contents to VW'.

3.2.4.1. System architecture

Despite the fact that VSS can run on native software, it was not the best approach for us because it would constrain the selection of libraries imported to the system. It entirely relies on third-party software. Blender, SketchUp, 3DS Max, Unity, Unreal Engine, Maya are tools for 3D modeling. We chose Blender as one of research colleagues knew already about this software.

Nevertheless, a suitable choice for us was to build the application on the web platform; by doing so, we could leverage many powerful open-sourced libraries in designing the virtual environment.

Figure 10 system architecture depicts a scenario in which a user creates an application using the Meshmoon rocket. The system handles resource upload as well as the establishment of the access route to the application using a webpage or the native software. For instance, it uploads all the necessary files to the Amazon cloud data storage at the time of creation of the application, which then stays running to serve the incoming requests from clients. The connection requires authentication of the user.

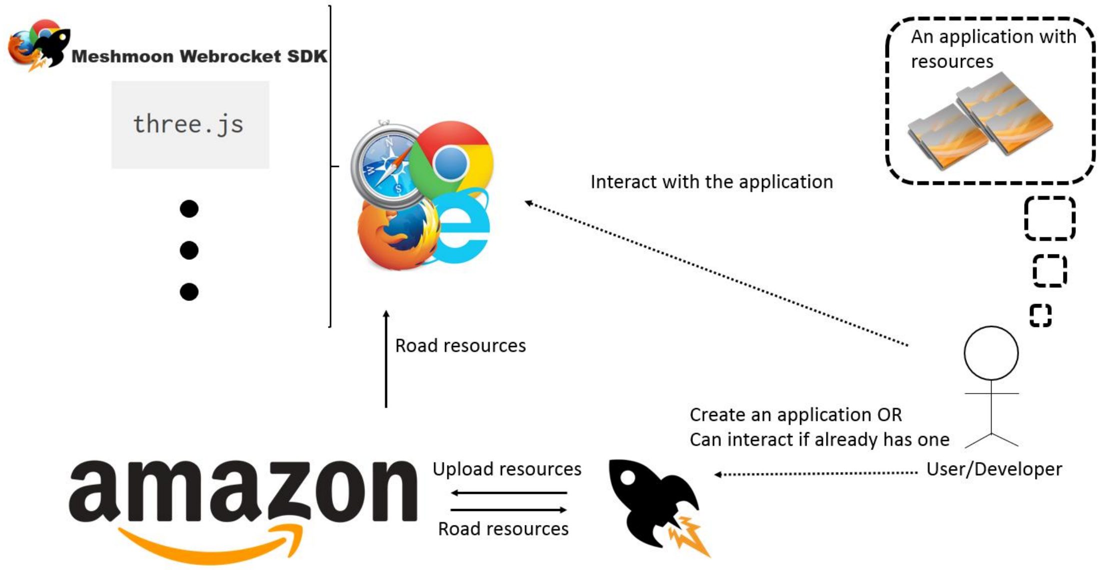


Figure 10. VSS architecture: the application runs on both the native software and the general web clients such as Chrome, Firefox, Internet Explorer, and Safari.

Looking at the application in Figure 11, it comprises of several files, each responsible for system components (server, client, and assets/resources). The Meshmoon web platform sits on the top layer of the system hierarchy, which receives all signals sent by client machines that are present in the application. It is also an access point through which the application uploads all resources to the archive; once uploaded, client machines can directly send HTTP requests to the Amazon storage centre, and fetch them without going through the server again.

The application layer contains material and mesh files which are exported from the 3D modelling software. The content of these resources specifies information such as, where the object is placed and which object to associate with. A mesh object typically needs coordinate information indicating where it locates in the 3D space, and a material file describes how the mesh to appear in the VE (the type of attributes include colour, depth, transparency, diffusion, light, and so forth). Also, the server system script works with incoming data from the clients, which it arranges, adjusts, synchronises the information, and distributes back in a uniformed format to all client connected.

Finally, the bottom of the hierarchy engages client systems with a 'webrocket' file; its primary role is to load the space components into the machine and reconstruct a VE. The file is executed on the client side as the user joins to the application. While developers are eligible to leverage open libraries to enhance graphical quality and visual capability of the VE, the server and client script inherit Tundra APIs used to enable 3D modelling. The APIs also provide means of easy handling for the low-level implementation of a VE such as, client log-in/-out, RPC-like signal communication between systems, graphical user interface for object modeling, and changing attributes.

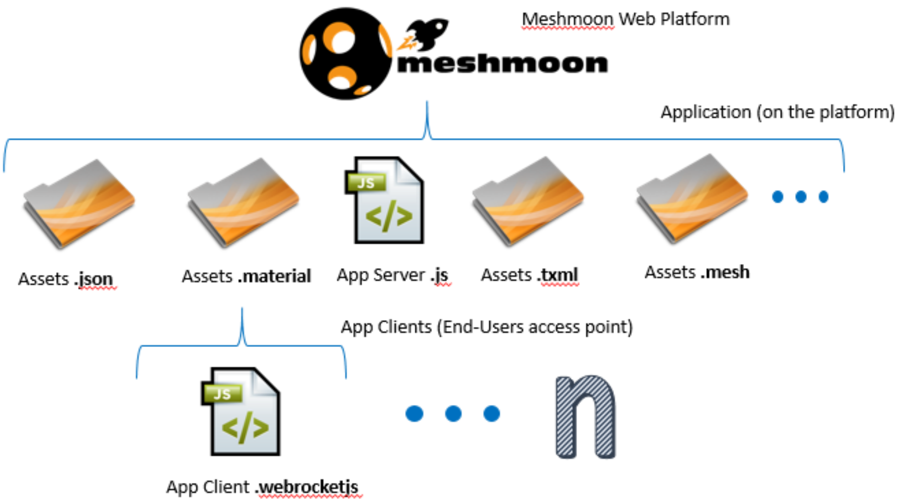


Figure 11. Hierarchical architecture design of a Meshmoon application, expanded into detail elements.

3.2.4.2. System requirements analysis

This section presents a second SORs of the system. While the first analysis examined the system at its interest for stakeholders, the second explores in-depth specification of the system behaviours. A use case diagram in Figure 12 will help understand the actions in the aspect of systems involved. Also, the second boiler-plated template is reviewed (‘update’ here donates general status changes such as modify, add, and remove).

1. The *<server>* shall be able to *<functional requirement> <non-functional or performance requirement>*
    a. The *<server>* shall be able to *<update client information> <if a client joins or leaves the system>*
    b. The *<server>* shall be able to *<update coordinator information> <if a new coordinator is elected>*
    c. The *<server>* shall be able to *<update objects information> <if an input subject to the object arrives from the client>*
    d. The *<server>* shall be able to *<update speech information> <if a client submits a new speech info>*
    e. The *<server>* shall be able to *<store speech information> <if a new speech information arrives>*
2. The *<client>* shall be able to *<functional requirement> <non-functional or performance requirement>*
    a. The *<client>* shall be able to *<store other client information> <if a new information arrives from the server>*
    b. The *<client>* shall be able to *<update objects information> <if a new information arrives from the server>*
    c. The *<client>* shall be able to *<update coordinator information> <if a new information arrives from the server>*
    d. The *<client>* shall be able to *<input command keys> <always>*

e. The <*client*> shall be able to <*update speech information*> <*if a new information arrives from the server*>

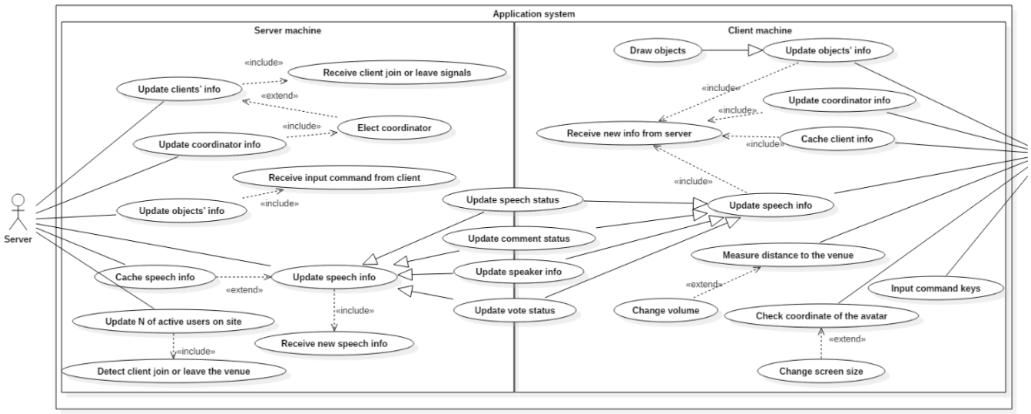


Figure 12. Virtual Soapbox System: System-centered use case diagram

Once again, the actions seen in the diagram are further taken into an analysis by using activity diagrams. It draws a top-to-down flow of the event. The activity diagram also allows us to point out in which particular step of a function involves communication between systems. Features covered here come down to; joining and leaving the VSS (Figure 13) and the process of input comments (Figure 14).

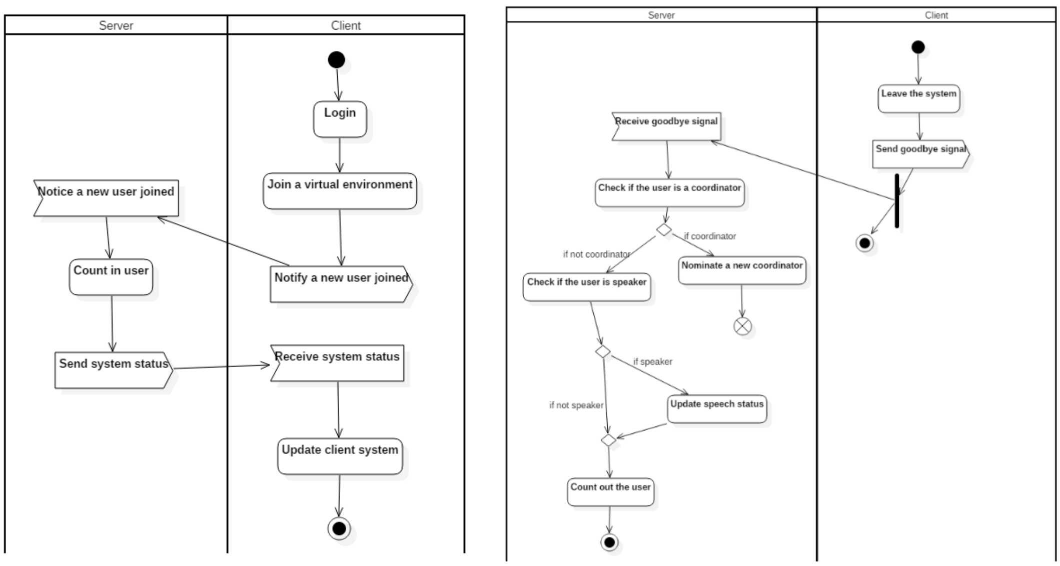


Figure 13. Activity diagrams: the left shows an internal action flow for the case which a client joins the server and the right; a user leaves the server.

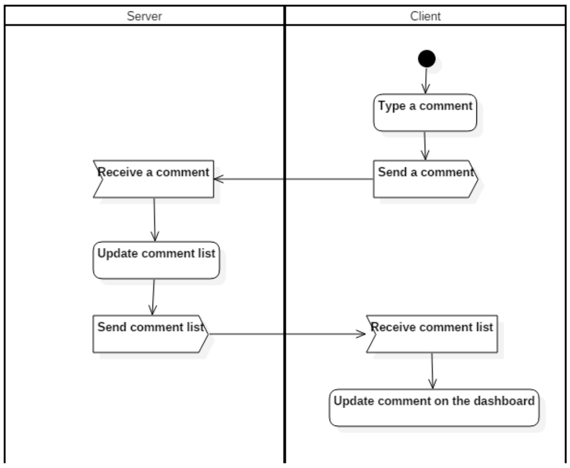


Figure 14. Activity diagram that shows how the writing comments is handled within the system

Down the lines, there have been no chapters that explain the role of ‘coordinator’ in the system (however, presented in previous figures); there is, however, a vital role of it. Because of the developmental convenience in creating a Meshmoon application (server side scripting is terribly user unfriendly), developers tend to build the system biased to the client side. For instance, upon our video transmission, VSS establishes the connection to Middleware directly from the user machine. That being said, we needed a logical algorithm to correct the inconsistent data within the VSSs, and that is where the 'coordinator' comes to play. VSS elects a user to get this job done, and keeps the information up-to-date in case of he/she leaves the system (select another coordinator).

3.3. System implementation

The implementation of VSS contributes to both server and client systems. Underlying network communication between them utilises UDP protocol sending signals. The low-level networking implementation is readily dispensed in the form of RPC-like callback methods from Tundra APIs—The technology is developed with an intent to ease the workload of low-level programmings, such as networking, behaviours for creating VW applications, and basic 3D modelling for the developers.

As seen in Figure 11, VSS structures generic one-to-many, two-tier server-client architecture design. The server side runs constantly, whereas the client script is executed on individual client computers when the user opens the application. The service provider of the platform provides the framework and hence, we do not need to worry about platform-level implementation.

3.3.1. Enabling technologies

We have built VSS on the Meshmoon platform. It allows developers to create customised VEs with their own objectives. The software application technology follows a two-tier client and server architecture which allows developers to modify server side script in comparison to the SL.

Entity-component approach that origins from game engine architecture along with generic mesh and attribute models for virtual applications allows developers manipulate and control the 3D models at hand.

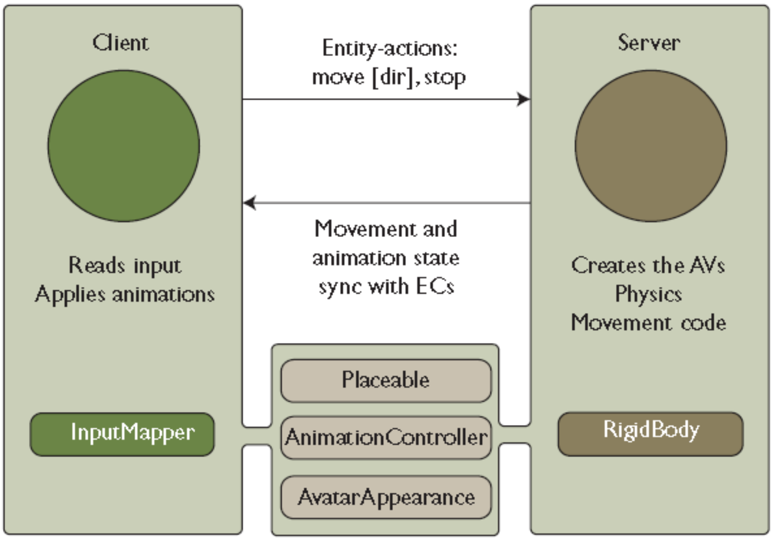


Figure 15. Conceptual illustration of an avatar example of entity-component model, from Alatalo (2011)

Figure 15 exemplifies how the entity-component system design functions for avatar creation and its control. In this particular example, the client is responsible for avatar visualisation, looks, and animations exclusively, whereas the server holds control over coordinate of the avatar and its physics attribute. Key and mouse inputs are mapped on control of motor ability of the avatar. The base form of these entities comprises of Mesh, Name, and Placeable components, for which each controls geometry, identity, and location of the entity. And the component contains many attributes referred to the local or remote resources that describe the entity. By way of example, let's say there is a 3D mug, what it decides the shape is a geometry attribute that is configured by Mesh component; a mug may have a handle and circle or square rim on the tip. Similarly, Name component has attributes that represent the identity of the mug; name, identifier (indicating to which group the entity belongs), description, and so forth. Placeable component contains the coordinate of the mug in the VE and entity visibility.

The entities used in a scene are composed of the txml file, which the term stands for associating an XML formatted, readable text file; this can accommodate information of multiple VWs for any clients to load. The content structure of the file is designed that is easy to read for human [82]. Figure 16 below shows an example file.

```xml
<entity id="10" sync="true">
 <component typeId="17" type="Mesh" sync="true">
  <attribute value="0,0,0,0,0,0,1,1,1" type="Transform" id="nodeTransformation" name="Transform"/>
  <attribute value="http://tundra-blueshift.s3.amazonaws.com/VirtualOulu/ubi2015/Asemakatu4.mesh" type="AssetReference" id="meshRef" name="Mesh ref"/>
  <attribute value="" type="AssetReference" id="skeletonRef" name="Skeleton ref"/>
  <attribute value="http://tundra-blueshift.s3.amazonaws.com/VirtualOulu/ubi2015/Asemakatu4.material" type="AssetReferenceList" id="materialRefs" name="Mesh materials"/>
  <attribute value="0" type="real" id="drawDistance" name="Draw distance"/>
  <attribute value="false" type="bool" id="castShadows" name="Cast shadows"/>
  <attribute value="false" type="bool" id="useInstancing" name="Use instancing"/>
 </component>
 <component typeId="26" type="Name" sync="true">
  <attribute value="Asemakatu4" type="string" id="name" name="Name"/>
  <attribute value="" type="string" id="description" name="Description"/>
  <attribute value="" type="string" id="group" name="Group"/>
 </component>
 <component typeId="20" type="Placeable" sync="true">
  <attribute value="-121.829216,7.24652815,73.1547699,0,-32.7589684,0,1,1,1" type="Transform" id="transform" name="Transform"/>
  <attribute value="false" type="bool" id="drawDebug" name="Show bounding box"/>
  <attribute value="true" type="bool" id="visible" name="Visible"/>
  <attribute value="1" type="int" id="selectionLayer" name="Selection layer"/>
  <attribute value="" type="EntityReference" id="parentRef" name="Parent entity ref"/>
  <attribute value="" type="string" id="parentBone" name="Parent bone name"/>
 </component>
</entity>
```

Figure 16. An example code for an entity representation in txml file

Nonetheless, one great feature that accounts from Meshmoon is that it allows users access VEs without any prerequisite installation from third-party software; This is possible through Tundra implemented into the web with the help of WebGL library. The attributes of high accessibility and availability of web-based VE application have drawn a great opportunity towards the VR and VW community. Also, nowadays researchers can easily involve publics using web-based 3D space.

The Meshmoon team bases in Oulu, our local city, which we see it beneficial for us, in the case of technical issues. On top of that, we have some prior knowledge of the Meshmoon technology. All in all, we decided to choose it for our application development platform over other competitor software. JavaScript was the only available option for the programming language on both client and server sides.

Threejs library powered up the creation of 3D objects for the application. Threejs is an open sourced library, one of the most popular WebGL frameworks that exist and has been gathering more contributors in the project since it was born in 2010 by Ricardo Cabello (a.k.a Mr.Doob, see [83] for details). Besides, jQuery, Bootstrap, Classy-js, Moment.js, Tock.js, and many more libraries are used for the application implementation.

3.3.2. Realization of the implementation

This chapter presents thorough report on system implementation. Features are acquired by the requirements analysis process which we covered in a twofold process in the earlier chapters. The first was devoted to the user-centered design which has generated a boiler-tailed statement of functional and non-functional requirements of the system. And followed by a second analysis with an aspect of more particular system task wise features. At the end of implementation, we have written six JavaScript files—contain six classes—for the client and server system which we will present in this chapter with the help of class diagrams, action flow diagrams, and subsequently sequence diagrams.

Four classes are closely associated with the operation of client system: two control classes and two boundary classes. The 'VirtualWebClient' is a main system class that responses to the inputs of the end-user, and is an output gateway for the other client classes. It is primarily responsible for receiving user inputs and sending the acknowledgement to the user accordingly; more context specifically, the corresponding behaviours include drawing 3D objects, resizing models, and adjusting attributes of 3D objects. Also, the class works for caching up-to-date information arrived on the client system; for instance, ‘initParams()’ method is being called at the time of loading the client viewer to configure initial states. The ‘Dashboard’ class, on the other hand, is in charge of web dashboard interface for the display of shared information. Once a 'speech-begin' request arrives at the client, it draws the dashboard UI on the client web view.

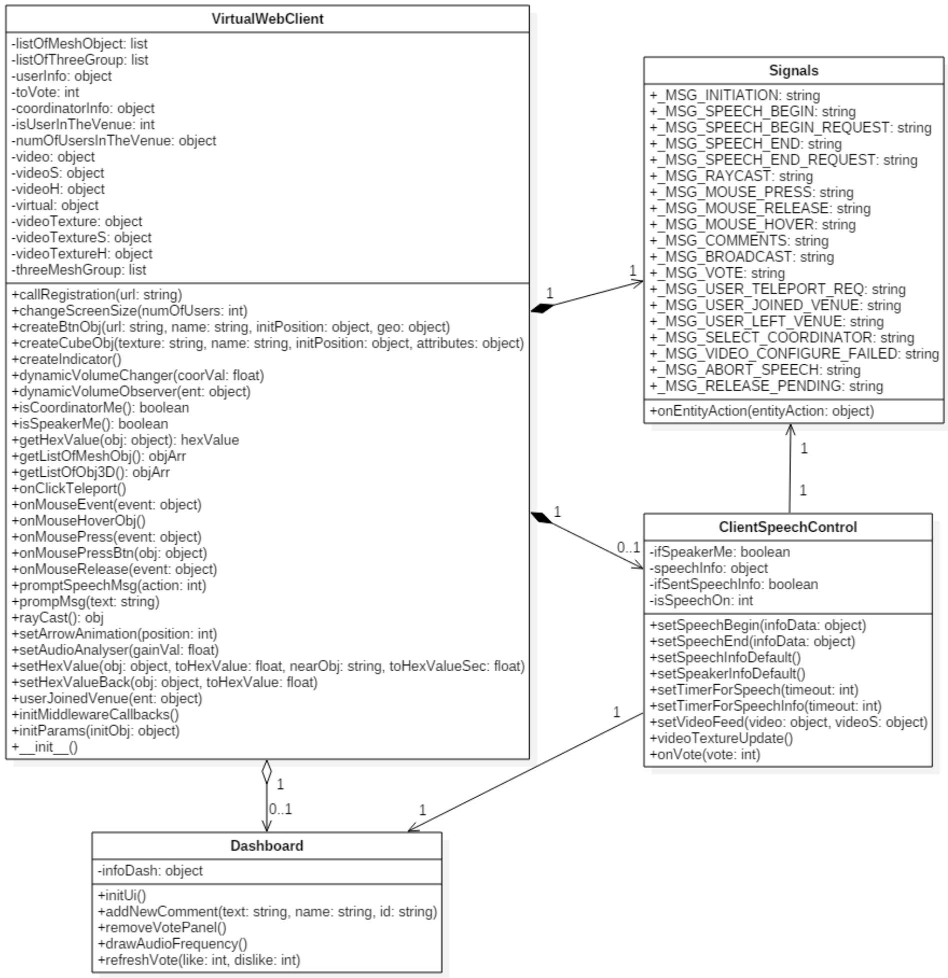


Figure 17. Class diagram: Virtual Soapbox client system

The rest of ‘Signals’ and ‘ClientSpeechController’ classes are the controllers in the system, for which each contributes to processing signals and coordinating speeches. When a client either attempts to reach the server or a signal arrives at the system, the Signal class cares the process. Lastly, ClientSpeechController class governs all the necessary functions for the speech activity, for instance, if the user attempts to give a vote, ‘VirtualWebClient’ notices it and executes ‘onVote()’ call to perform the request.

Moving on to the server system classes, we have two classes in the server system; ‘VirtualWebServer’ (VWS) and ‘Users’. The number of classes is comparatively small, in the result, VWS remains as a single master class controlling all the functions of the server by itself except maintaining client information. The VWS class stays running on the Meshmoon web platform once the application is initiated (uploaded). It governs communication task with new clients joining the VE. VWS also uses 'Users' class to manage the ledger for the user information, such as query history, request detail, IDs, and avatar information. For example, if a new client enrols to the server, 'Users' class receives the primary signal and relays it to VWS class after registering the user in the ledger.

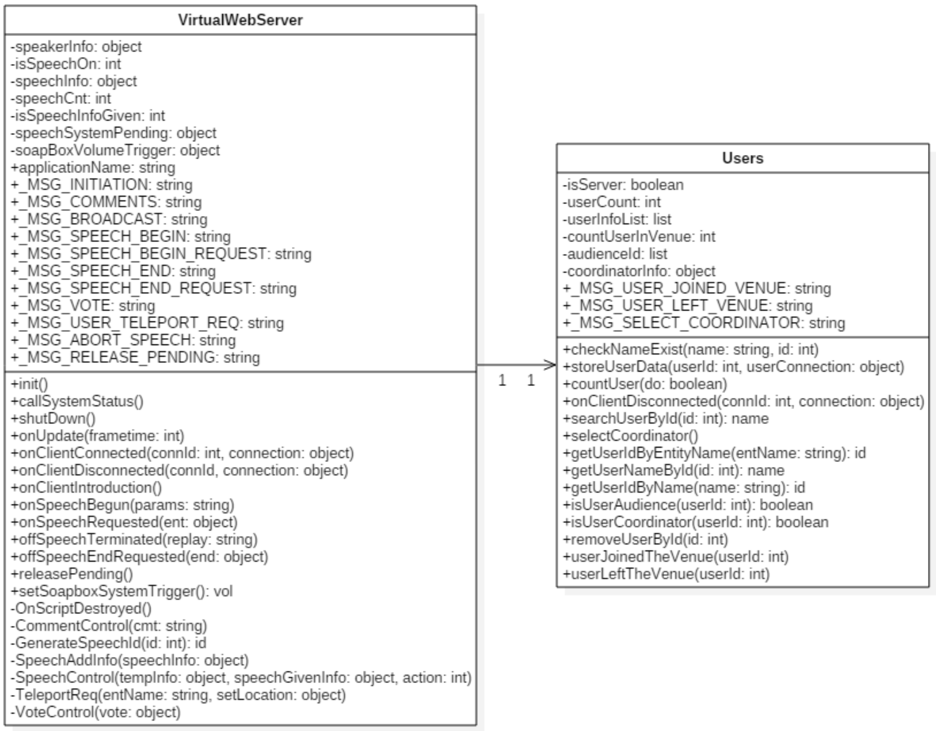
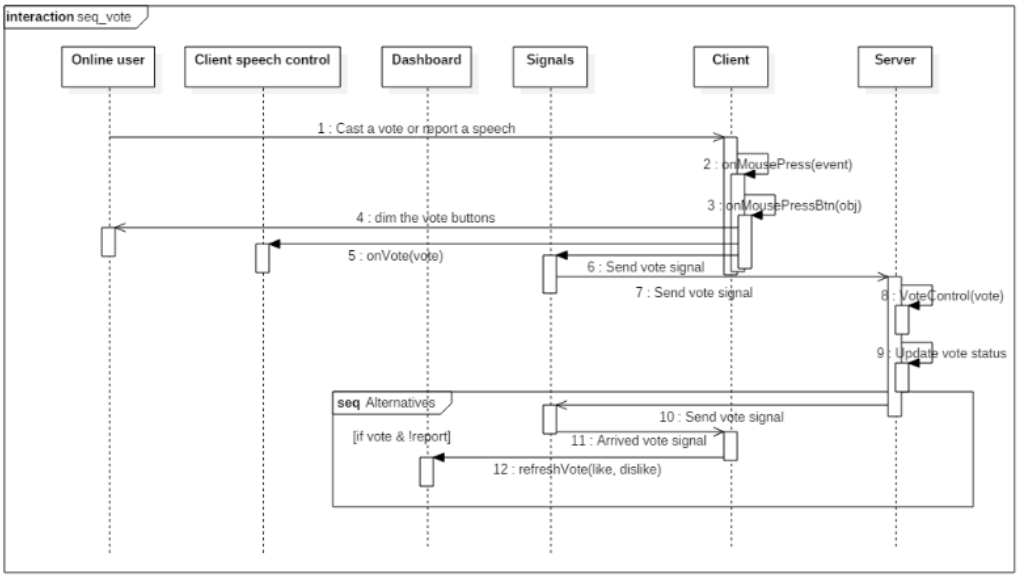


Figure 18. Class diagram: Virtual Soapbox server system

3.3.2.1. Main features for speech activity

Online users can begin and end a speech as well as other speech-related activities, such as voting and commenting. When a user first time joins the application, it spawns an avatar at the pre-configured coordinate near the speech venue in the virtual space. If there is a speech on-going, the user sees virtual screens standing by the spawn point streaming the live view of the speaker as well as the view of real environment captured by a camera attached to the hotspot—a public display that broadcasts the speech; also, used for audience to interact. The box buttons are highlighted as a mouse cursor hovers over, indicating that the user is allowed to give a vote, otherwise the button is dimmed until the next speech begins. In the meanwhile, the user would be able to see the dashboard on the right side of the web view, in which displays information, such as remaining time of the speech, vote status, and so forth. Also, the comments are seen by both audience and speaker, making the dashboard interface a communication window.

In case there was no on-going speech, everyone has the right to start a new speech by stepping on the virtual soapbox; it should be obvious for anyone how the soapbox is used as a red arrow pointing down moving above the box. Once a user steps on, the system shows the dashboard asking to enter speech information with a timeout. Unless the user leaves or does not do anything until the timeout, the system will forward the complete information to the application server as well as the Middleware system requesting a new speech begin. Furthermore, we implemented on the dashboard, a voice feedback feature, which supports a user experience; it draws a graphical feedback in response to the voice input. And a volume modifier to adjust the speech volume helps speaker easily control the sound output.



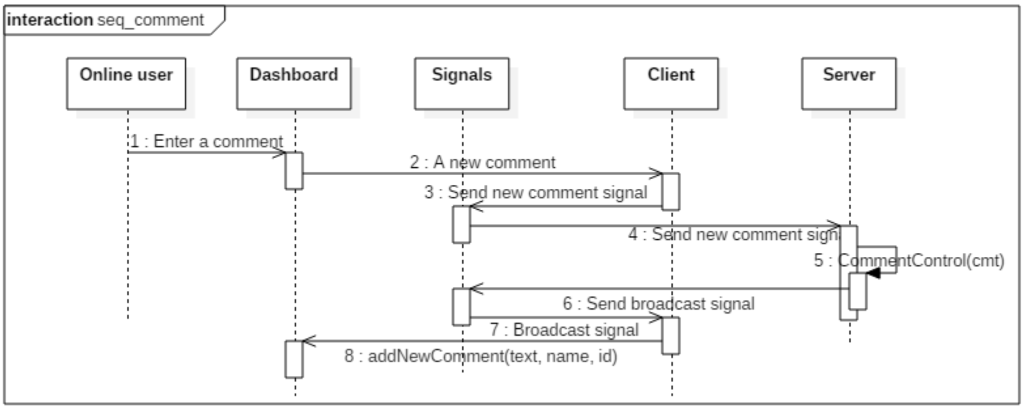


Figure 19. Sequence diagrams: an online user attempts to give a vote (upper); to add a new comment (lower)

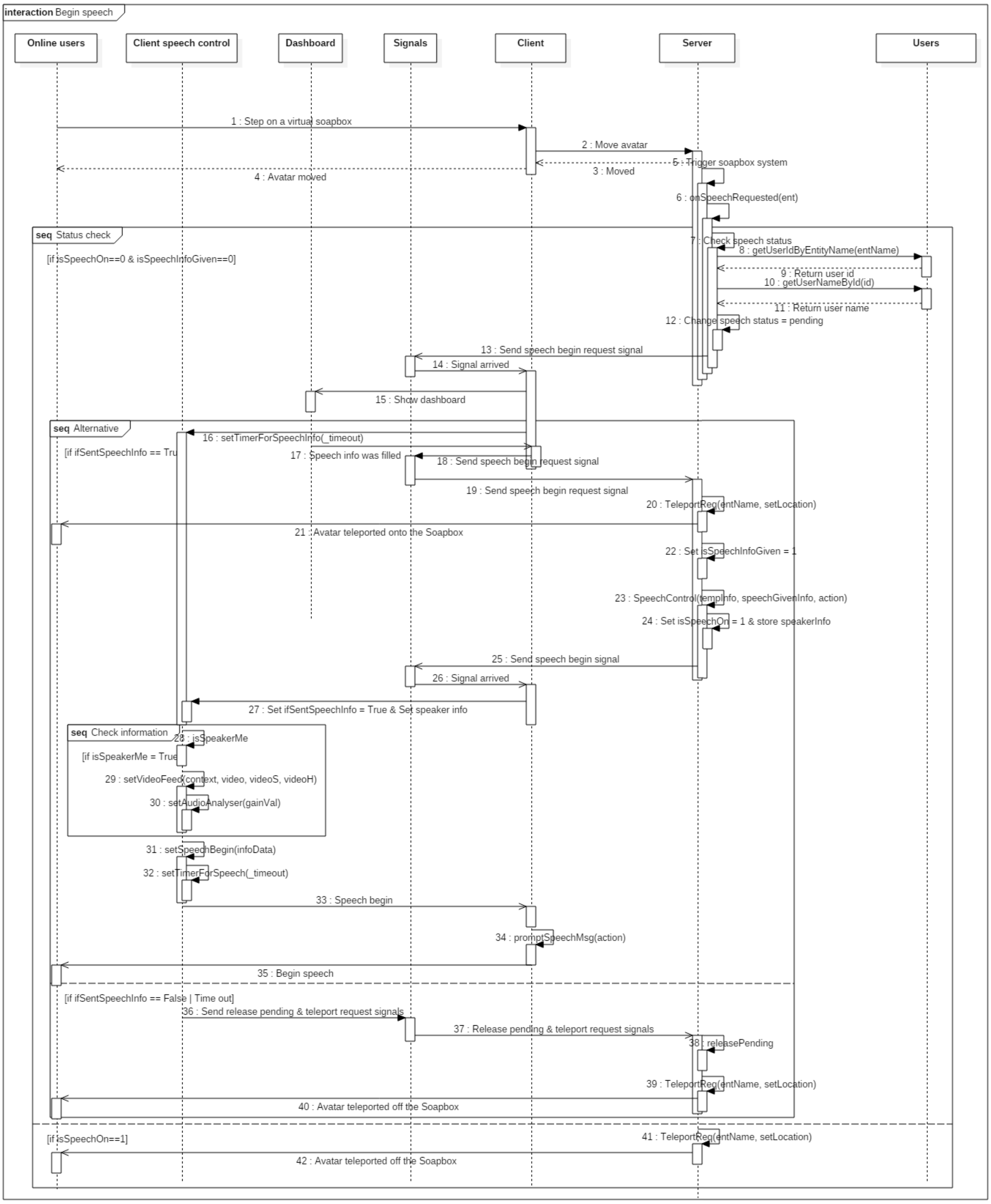


Figure 20. Sequence diagram: an online user attempts to begin a speech

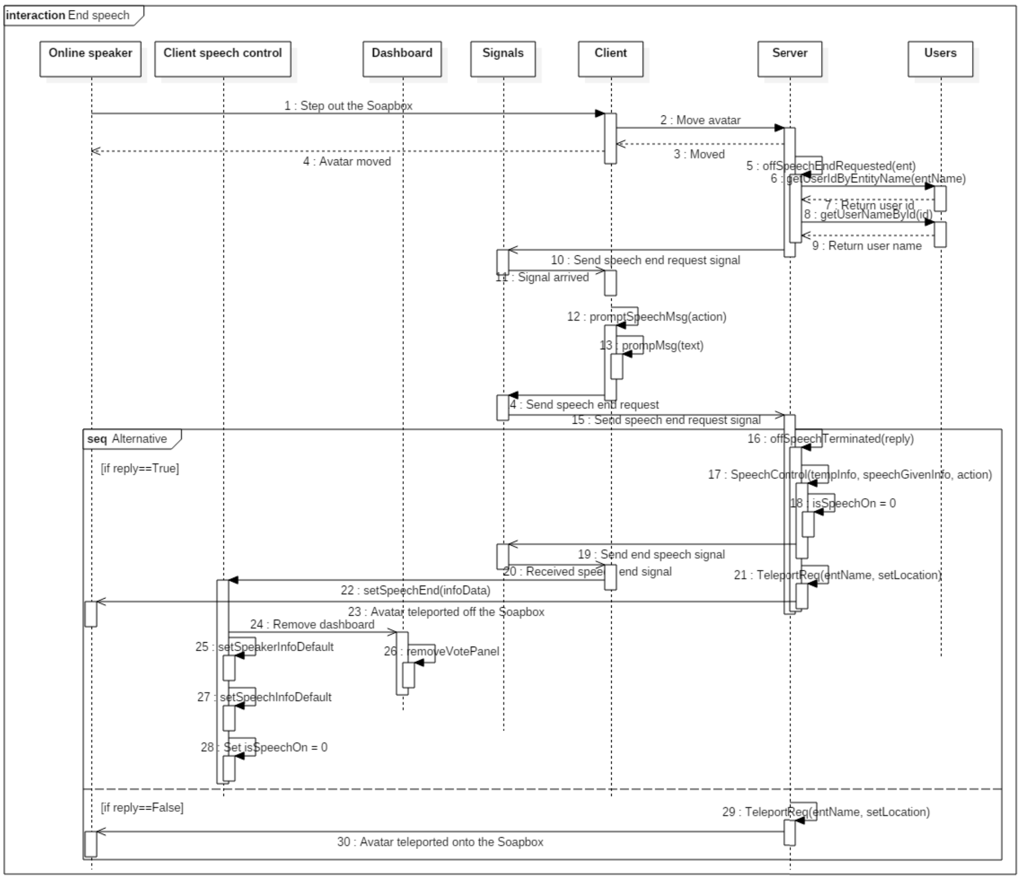


Figure 21. Sequence diagram: an online speaker attempts to end a speech

3.3.2.2. Additional features in the system

Some of the additional features added throughout the iterative development process are already shortly mentioned in the previous chapters. Here we will discuss what these features are and why added to the system.

Firstly, we added feature what we call (a) ‘dynamic scaling of the virtual displays’. It is a controller that surveys the number of online users currently standing within the boundary of the venue. Such information is useful for scaling the display size according to the user population. It ensures visibility of the contents in the screen under the circumstances which there are less or more people. The feature is realised in the client system mostly, with a partial cooperation of the server system.

The algorithm behind the feature investigates if there are more users in the venue than the threshold value. Nonetheless, the laborious aspect of the operation is locating the coordinate of the avatar; the client machine has to check this information in the main loop which is an arduous duty for the hardware concerning throughput.

The next feature added to the system is so-called ’dynamic volume changer’. It is an observer designed to provide a real auditory experience to the online users in a virtual space. Similarly, the synchronised audio feedback and sound effect were invoked in the literature as one of the prominent sensory perceptions of VR system at its goal for maximising the matching between proprioception and sensory feedback at the perceptual and cognitive level, referred via terms ‘Matching’/’Mapping’ [33]

[46]. The logic behind the algorithm lies in the varying coordinate of the avatars. The system predefines the area of different zones by ten levels (0 with highest volume scale and 10 with muted to 0 volume, see Figure 22). As the avatar walks across zones, the volume observer reports the information to the relevant system, and then its controller adjusts volume scale accordingly.

The algorithm that investigates the particular zone in which the avatar is currently standing can be presented mathematically as $A \rightarrow Z_A, B \rightarrow Z_B$ where $A = |x_c - x_a| \geq |y_c - y_a|$ *and* $B = |x_c - x_a| < |y_c + y_a|$, such that $Z_A = \left\lfloor \frac{|x_c - x_a|}{D} \right\rfloor + 1,\ Z_B = \left\lfloor \frac{|y_c - y_a|}{D} \right\rfloor + 1$. x and y donate coordinate of the center point of the lowest zone as (x_c, y_c), where the coordinate of avatar is (x_a, y_a). D donates the distance between two adjacent zones wherein $Z = \{z_0, z_1, z_2, \ldots, z_n\}$ is a set of auditory zones in the virtual space. Therefore, A and B calculate the comparative two-dimensional distance between the center point of the lowest zone and the coordinate of the avatar, such that, Z_A *or* Z_B is the zone where the avatar belongs in. If Z_A *or* Z_B is larger than z_n, then the result becomes z_n.

The algorithm works well also for the different schema of the zones (i.e. circular boundaries), in which case, we do not need to consider separate conditions, such as A and B instead, we can directly apply $Z = \left\lfloor \frac{|x_c - x_a|}{D} \right\rfloor + 1$ or $Z = \left\lfloor \frac{|y_c - y_a|}{D} \right\rfloor + 1$ to locate the zone where Z donates the target zone in which the avatar is standing, however, the rest of conditions remain unchanged.

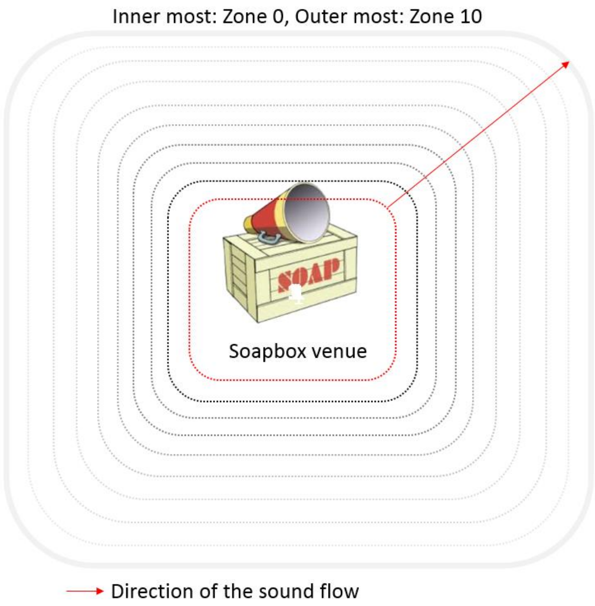


Figure 22. Dynamic volume changer zones; zone 10 to 0 volume, zone 0 to maximum volume

The system operation for the task runs entirely within the client system, and hence, the server does not know about the volume status of online users at all. Again, the throughput of the feature is substantial as the system regularly calls necessary functions inside the system loop.

Nevertheless, we expect this feature along with the previous one can raise interesting topics for the research on the human behaviour in VEs. Particularly the latter feature would be able to support inducing a natural and realistic experience of VEs in the auditory aspect.

4. EVALUATION

As the purpose of this study was to evaluate the influence of immersion and presence levels of the user to performing a free speech in a virtual environment, we ran consecutive six semi-controlled lab studies with six participants recruited through public advertisements using posters and direct email contact. Each experiment was designed to allow the speaker to give a free speech to the audience. Throughout the speech, we expect the audience to imagine the speaker be located in virtual space, represented as an avatar by showing them side by side live view and virtual view—one the real-view of the speaker in the CAVE and the other the avatar view in virtual space. Audience could send comments using a web application that can be accessed by scanning the QR code found in the hotspot interface. Quantifying immersion and presence levels are done with Witmer and Singer' questionnaire [5] consisting of twenty-four questions for each (measured with seven points scale). In each experiment, we focus on the interaction factor with the audience that will generate feedback-like data upon which we can evaluate the quality of performance of the speaker. Across the six experiments, we made a number of changes to the experimental configuration, which we title as field trial one and two in the following chapters.

4.1. Experimental settings and participants

The experiment took place on the campus of the University of Oulu. Participants were invited to the CAVE system; it is a physical space designed to build an immersive virtual reality environment, equipped with three projectors facing the wall of a room-sized cube. Such artificial environment with various VR hardware devices can provide a user with realistic experience inside space. In our study, we do not include VR components such as head-mounted displays. Instead, conventional input devices such as keyboard and mouse are used in the virtual space.

On the other side of the campus, a large public display (57-inch LCD panel) sits there for the purpose of broadcasting the speech, and it works as a medium for the communication between speaker and audience. When the experiment initiates, the system transmits view and voice of the speaker onto the public display, so that passersby that may find it interesting can view or interact. The public display has a web user interface designed to accommodate two video views, comments section, QR code for the mobile application, and three buttons for voting. When a user scans the QR code, it opens a mobile web application through which audience can vote and send comments after entering basic information such as, gender and age. Vote status, as well as comments history, are displayed on both interfaces of speaker and audience (see Figure 23).

Concerning the CAVE UI, we spared an extra room for showing comments and time remaining of the speech by replicating the contents onto multiple faces of the wall. It eases the speaker capture information instantly.

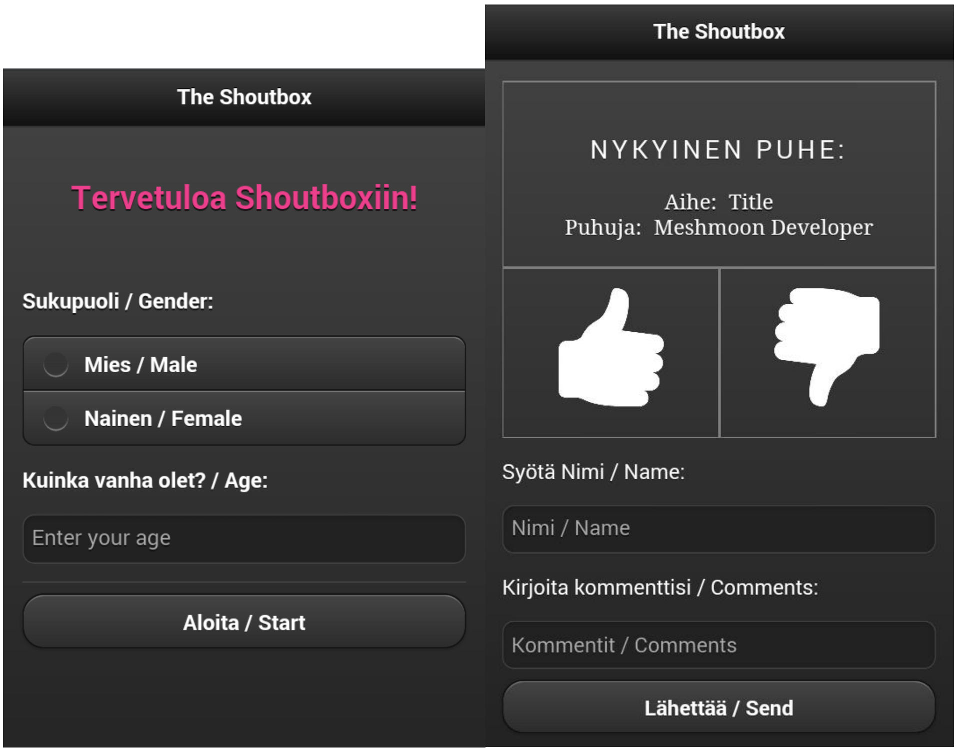


Figure 23. Mobile web application user interface

Participants were six individual students and researchers from the department of computer science whose age range between 22 to 28. Gender distribution was not regarded as a variable since there was no literature mentioning that it affects the immersion and presence of the user. Therefore, we carried out with six male participants. Some of them preferred to bring a script to read from throughout the speech. However, most of them chose to speak improvisationally. We will cover these cases separately in a later chapter.

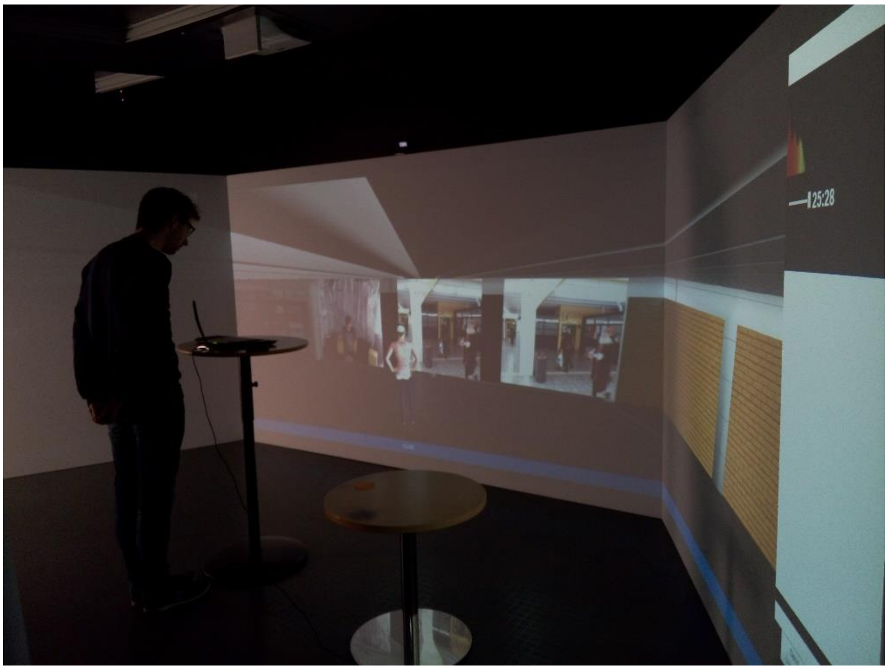


Figure 24. Participant giving a speech in the CAVE

4.2. Field trials

4.2.1. Variables

We assume the spoken language of the speaker would certainly influence the experiment results. Thus, we constrained the speech-language option as English. We expect that everyone on campus understand English. A topic for the speech is also a critical independent variable which can lead to a biased opinion of the audience. For instance, some people find a specific subject more interesting than the other therefore an individual's preference onto the contents gets to play a bigger role than the other attributes. To this end, we predefined a theme of the speech as a university or school-related subject for all participants. The selection of the place to install the hotspot can also influence the research outcome; the more people the system is exposed to the higher chance it gets to entice people. Speech volume and illumination inside CAVE are also considered as independent variables that are particularly associated with the user experience; we tried to keep the sound output consistently maximum and installed a stage lighting in the CAVE that would help improving visibility of the speaker to the audience. Experiment date, timing, and speech length are also important variables too, for which each has a possible influence to result; In our study, we define the date and timing as independent and speech length as dependent. Weekdays (Wednesday and Thursday), busy hours, and longer speech duration were preferred. To the better generalization on these independent variables, we ran a series of observation sessions conducted on the population measure around the display (Figure 25). Lastly, we are assuming that the presentation style of a speaker may affect the ability of the audience to focus on the speech. Therefore, we left it to the participant choosing the method to present their speech.

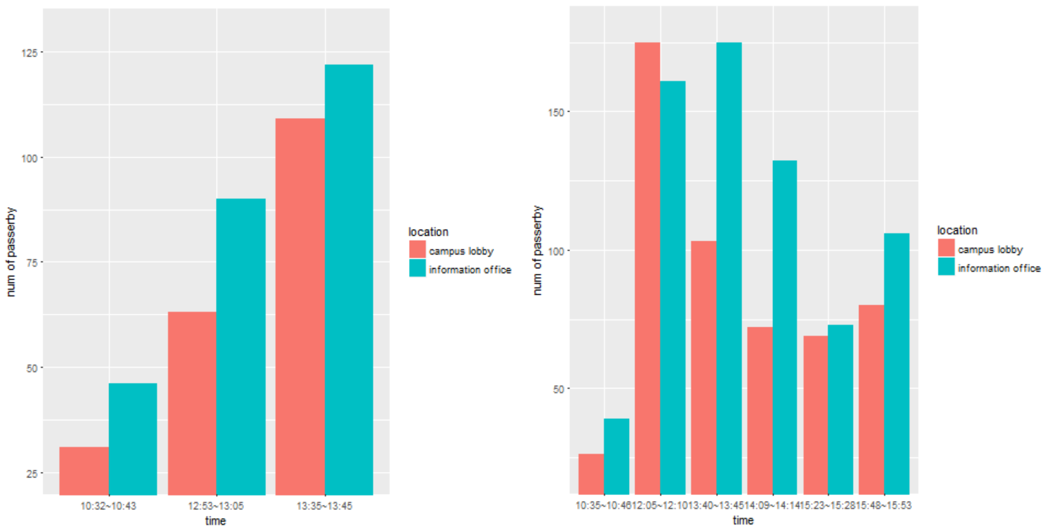


Figure 25. Headcount observation on two different locations (indicated red and blue) for the hotspot (left: Wednesdays, right: Thursdays).

4.2.2. Design and procedure

Each participant was briefly introduced to the system design and the study purposes before entering the CAVE. While the participant was asked to explore the virtual space, we set up the other systems. The selected locations for the hotspot were based on crowdedness, one by the main lobby of the campus and the other near the information centre (both places are interchanges of many paths). It is also the scene of the virtual world shown in the display (Figure 26).

As the hotspot sets to begin, we cued the speaker to start the speech. During the experiment period, we approached some of the people that stopped by the hotspot or used the mobile application to interact with the system, and asked to fill out a survey designed for the audience (which for the later experiments, switched to an interview method as it turns out filling a survey was such a hassle to them). To avoid distraction, we approached audience members when only they were about to leave.

Figure 26. Public display in 'waiting for a speech' mode showing a pseudo-avatar in the display

4.2.3. Measures

We collected data through two methods: on the speaker side, we asked to fill out a set of questionnaires to measure the sense of presence and immersion, and on the audience side, we used observation and interview methodologies. The contents of audience interview consist of information as follows:

1. Basic information such as age, gender, background and major.
2. Motivation for 'joining' the experiment.
3. Usage of QR code / possession of QR code scanner.
4. Opinion about the speech content.
5. Any experienced technical failure such as low sound volume or system dysfunction.

6. Opinion about the virtual world view alongside the real-world view shown in the display.

4.3. Changes to experimental settings

Over the early experiments, we encountered few technical defects on the hardware, which hindered us from carrying on the field test. Hence, we took off some days, resolved the problems. Similarly, experiment after another, we improved the research design and environmental settings based on feedbacks received from people we interviewed and also our observation. The followings report the list of changes we made both hardware- and design- wise:

1. Moving location where we deployed the hotspot system:

Two places are apart each other about two minutes walking distance.

2. Enhanced sound of the speech, and the better illumination inside the CAVE:

In the early test runs, we had a difficulty dealing with loud noise coming from crowds. Then we tried an external speaker but eventually, switching the display hardware resolved the issue (the speech could be heard from 15 steps away). Also, the real view of the speaker was dim that one could not clearly see the person in the CAVE, so we placed a stage light in the room in such a way that it does not interrupt the speaker while ensuring a clean view to the audience.

3. Active support for the interaction with the speaker:

In hope for more users to engage with the experiment, we added QR code and made them more visible to people. Later on, we also provided 'ready-to-play' mobile phones which have already mobile application installed.

4. Data collection method.

Asking to fill out the survey paper to audiences likely burdens on them with more works, hence, we interviewed instead based on the survey questions, which turned out people become more responsive in giving their opinions.

Figure 27. System field testing: the hotspot deployed near the ‘Information office’. People stop and interact with the display.

5. RESULTS

5.1. Observation and audience interview

We conducted in total seven field trials involving one recruited participant per round. Each experiment lasted 30 minutes to 1 hour with a 6~23 minutes of speech duration. The experiments were held in the CAVE room as well as two pre-selected sites on the school campus, chosen based on the crowdedness and the resemblance to that of the virtual environment. The speakers were invited to the CAVE to give their speech while the observation was carried out in the hotspot area for recording the experiment and interviewing audiences.

5.1.1. Results from field trials: Lessons learnt by doing

5.1.2. Learning phase

The first trial took place at 10:30 morning with planet of students around. However, we had to interrupt the experiment in the mid-way because there was a non-negligible noise coming from a booth few meters away from the hotspot advertising for their events. Speaker did not finish his speech, though we could sense some of the people seemed afraid to get involved with the system despite their curiosity. We see that such form of social embarrassment is a clear barrier for the future experiments.

A few hours later, we conducted the second trial with the same participant. This time, we added an external speaker to the hotspot to amplify the speech volume. There was still loud music near to us, but we requested them to lower the sound during our experiment.

Unfortunately, the new speaker did not improve the situation. The sound was still small and that later we found out that there was an issue with compatibility problem between hardware devices.

We decided to let the speaker finish the experiment, and in the meanwhile, interviewed two students that stopped for the hotspot. They came up to the display out of curiosity and stated that the system shown in the hotspot seems interesting. Of the overall experience were bad especially in the respect to the low volume setting which they marked that they did not hear at all. Also, we had low visibility of the speaker due to the closed environmental setting of CAVE, Also, we switched to interview method from questionnaire filling for collecting data as we felt it was very difficult to make people answer the questions of our interest.

One-third of people passing by appears to show a passive interest to the hotspot. We thought bothering audience to scan the QR code for interacting may build another layer of an invisible barrier to engaging the system; some people stopped to watch the experiment but soon turned back or continued to walk since our system was missing with sound component.

5.1.3. Field trials

The second participant joined the experiment early in the morning, 10:30, few days later since we had first and second runs. According to our observation on the population measure (Figure 25) on the campus, the amount of passer-by is relatively small during this period of a day. Sound issue was fixed. As it was still in the morning, there seem to be more people rushing to places than that having some other purposes such as waiting for someone, going for a meal and so on.

In total seven passers-by stopped for the speech and of two to three have spent more than a minute of time listening to our speaker. However, no actual attempts were made for the interaction. We speculated that the circumstance resulted might have been affected by the visibility of QR code adhered to the display (in addition to a small image in the hotspot UI).

The third participant carried on the experiment in the noon, same day. From this run, we confirmed that many average students in the university do not have an application for scanning the QR code. Our speaker found himself some audiences (total number of seven people spent minutes of their time in front of the hotspot), though they were not equipped with required tools to install our mobile web application. Interviews with two audiences also revealed a language regarded issue; of both responded that the speaker's pronunciation of English was difficult to follow. Though the bright side was, interviewees found the system interesting and were able to assimilate the virtual avatar to our speaker, except the fact that the design difference of avatar with the actual look of speaker was preventing from fully immersing.

The fourth participant took the turn a few hours after, 15:00, at which indicated having relatively small number of people around the campus. Based on the lessons learnt throughout the relay of experiments, this time, we prepared 'ready-to-use' mobile phones to provide people that want interaction with the display so, there won't be a hassle to download the application. However, together with the second participant's case, scarcity in passer-by resulted in low-frequency of interaction with people and thus, less data collection. We confirmed that the assumption of the relation between the number of passer-by and chance of interaction is valid. These two runs were marked as the only rounds with no engagement at all with audience. Also, the timeframe of the speech was comparatively short, which might have been one of the attributes that deteriorated the consequence. Nonetheless, we observed that most people still gazed to the display on their move which is a positive sign.

The next field test with the fifth participant was arranged a week after, on the same date. Still early afternoon 13:00, there were more crowds than usual on the site, and a choir was singing accompanied by a piano. It was not terribly distractive though, loud enough to be heard from the area where the hotspot was sitting. The overly crowded place was filled with babbling and singing in chorus. As a result, we were able to observe only three passers-by stood by the display during the experiment period with no further engagement. One interviewee that we managed to speak to stated that it was merely by chance he was standing near the display and heard the voice. Although he failed to receive the idea which the speaker was trying to convey (speaking too fast and irrelevant topic how the interviewee describes), saw the whole system and how it presents a real person in the virtual space novel.

Our last participant begun the experiment about right after the previous one got done. The speaker did not have much to say in words though, strived for getting attentions using question phrases, body actions and gestures to the people passing by. Once the speaker encountered a group of own fellows approach toward the screen, he developed the positive atmosphere into a group action. At first, these audiences did not bother to use the mobile application for interaction, which then the speaker induced to utilise hand gestures from them for the conversation. As the time elapses with each other, as the more crowds gathering around the display, people started to use the mobile phone, sending messages to the speaker for a better quality of communication. We interviewed five actively engaged audience members, despite of the unique impressions towards the experience with the system for the individuals, almost common response was how the system works and presents to publics is interesting. The detail comments from the interviewees are presented in appendix 1.

5.2. Speech performance evaluation

The speakers are scored for the sense of immersion and presence in giving speech at the virtual space. For that, we use a set of questionnaires [10] to measure the experience quantitatively. Based on the field trial results (presented in table 2 and 3), we assess the performance of each speech and relate them to immersion and presence levels of the speakers.

5.2.1. Quantitative measure of presence and immersion

The sense of presence and immersion consist of multiple components which each influences the overall sensation. We employed the list of the sensory variable that was also addressed by Telepresence and Teleoperation to measure the speaker's feeling: 'involvement/control', 'natural', 'resolution', 'UI quality', 'control factor', 'sensory factor', 'realism factor', and 'distraction factor' are measured for the presence; and 'involvement', 'focus', 'games' for the immersion.
By contrast to the presence, immersion includes subjects of general experience in doing things. For instance, 'games' variable measures the tendency to play video games, and 'focus' to maintain focus on current activities.

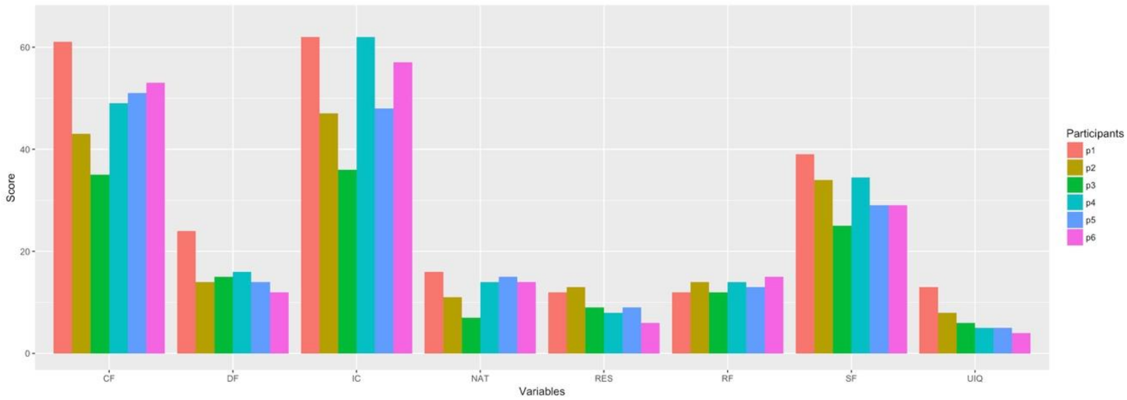


Figure 28. Presence score comparison between experiment subjects by system variables: CF=Control Factors, DF=Distraction Factors, IC=Involvement/Control, NAT=Natural, RES=Resolution, RF=Realism Factors, SF=Sensory Factors, UIQ=User Interface Quality.

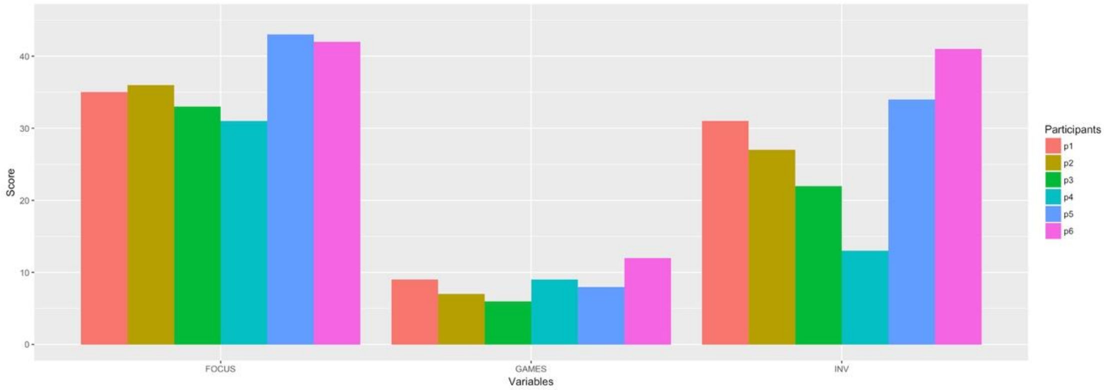


Figure 29. Immersion score comparison between experiment subjects by system variables: INV=Involvement, GAMES=Tendency to play video games, FOCUS=Tendency to maintain focus on current activities.

5.2.2. Performance analysis

In our series of experiments conducted to measure human sensation that occurs while being in a VE, we verified that presence of the subject is closely related to the his/her immersion level as noted by [10][59]. The Pearson correlation coefficient test shows the result is statistically significant by $r=0.97$ $(p=0.0011, <0.01)$. However, we failed to observe the correlation between length of exposure and presence on the basis of our experiment result $(p=0.773)$ [78].

In addition to the immersion and presence scores for the participants, we have collected data from audiences with which we could determine the 'quality' of the task performance (giving a speech). First of all, we need to define what the superior performance is as opposed to the inferior one. With this regard, since the vocal, phrase based interview samples constitute the majority part of the audience data, we decide to exclude them from the metrics for the performance assessment, and only take into account the numeric subjects which are active/passive interaction, comment/vote counts, and comprehensible score of the speech. The decision also attributes to the fact revealed during the audience interviews that the majority of interviewees experienced difficulties in understanding the speech content due to the various reasons. Nevertheless, despite the fact that those data are in numeric form, we still cannot include all because some of them are constant zeros across participants (i.e. comment and vote counts); therefore, we finally confirm 'passive engagement counts' to take for evaluating the speech performance.

Variables	**p-value**	**r**
IC	0.738	0.175
NAT	0.787	0.142
RES	0.437	0.395
UIQ	*0.047	0.815
CF	0.772	0.152
SF	0.437	0.395
RF	0.329	-0.484
DF	0.102	0.726

Presence total score	0.81	0.127

Table 1. Pearson correlation and coefficient tests on the sub-items of presence and the inactive engagement counts; *p < 0.05.

Variables	**p-value**	**r**
INV	0.453	0.383
FOCUS	0.937	-0.042
GAMES	0.831	0.112
Immersion total score	0.64	0.244

Table 2. Pearson correlation and coefficient tests on the sub-items of immersion and the inactive engagement counts.

The correlation coefficient test indicates no significant relation between total presence and immersion scores of the speaker and inactive engagement counts of the audience pertain, resulting in accepts our null hypothesis which presence and immersion have no influence on the speech performance. Subsequently, we carried on by breaking down the total scores into subjects—system variables—and run the same correlation tests to see any relation with the speech performance. The results still do not reject the null hypothesis except for the UIQ variable (Table 1 and 2), which implies that the interface of VE shows a positive relation with giving a high-quality speech. In other words, participants who liked the interface tend to perform better than those who did not.

We also suggested before that the speech duration has a close relation with the number of spectators one has achieved. To prove that, we ran a test which the result showed a strong linear connection between them *($r=0.921$, $p=0.009$, <0.01)*; this indicates that a speaker should have to perform longer in order to attract more people listening to the speech.

6. DISCUSSION AND LIMITATIONS

VW have been increasingly pervading into our everyday life. Some examples are applications being utilised for educational activities, social experiments, and simulations [61][38] [54]. To response the growing use of VWs, in this work, we explore a sociological experiment regarding the task performance made in a VE. We measure the phycological sensation of the users in association with its impact on the performance of giving a speech to the arbitrary audience using CAVE system and public display, where the presence and immersion are known to affect the user experience when being in a VE or using VR applications [75] [66].

6.1. Participant recruitment

In our semi-controlled experiments, we invite six participants to join for speech performance measurement. Albeit the system's primary target group was those people with a frequent chance of giving public talks and hence wish to practice by using our system, we could not make it happen because of very specific requirements and moreover, the lack of candidates. Instead, we focused on the secondary group and opened the call for everyone to join, which resulted in getting six individuals.

6.2. Speech performance measure

Measuring performance of giving a speech seems obscure, which it also accounts for many aspects to consider, and often, the threshold for selecting a good speech is different across individuals. Among many possible interpretations of defining a good speech, we deliver own rule relating to the number of spectators that each session of speech achieves. In other words, we determine the better performance on the given task by measuring how many passersby were enticed to the play. While the ultimate goal of this study is to see whether there is an impact of a psychological sensation of the human being towards a varying quality of the work done in a VE, we were not able to find a significant connection between them as presented in Table 1 and 2. The consequence might be due to the lack of sample data or unreliable experimental settings inevitably caused by the nature of field deployment, an inconsistent number of population on the campus can be an example. Also, we expect that the significance of the experiment would become greater by improving the selection of the participants; for instance, inviting the primary users group of the system.

6.3. Audience interview

During interviews, the majority of interviewees raised an understanding issue (6 out of 10). Whether or not that was the result of lack of concentration, or speaker's fast/vague speech, or poor sound quality of the display, the consequence made it difficult for the audience to assess the speech by the given content.

Nevertheless, we present a summary of interview results here to provide an insight of the overall atmosphere of the experiment.

In total, we interviewed ten audience members with questions asking general impression of the system as well as the 'quality' of the speech; two of our participants did not have the audience at all.

First to note from the data is the novelty of the installation; seven out of ten interviewees stated that the system is interesting and new to them which have become the reason for spending their time. This could correspond to our initial assumption for the experiment condition that the number of people the system is exposed to can be a crucial factor which has a strong influence on the inactive engagement counts because a high proportion of audiences come from curiosity at first. However, unfortunately, we do not have the control over the university dwellers, and hence, we did our best to operate all experiments under even environment possible by referring the population graph in Figure 25.

The second point that grabbed our attention was the lack of QR code scanner on the mobile phones; eight out of ten audiences did not possess the application. One person specifically addressed that he does not like the idea of having to use a third-party application to communicate, and another person also indicated that it is not convenient interacting through the mobile phone while facing with the speaker in the display, would rather prefer voice chat approach. However, there is a tradeoff between two communication methodologies; commenting based design allows more audience can engage with the speaker without having to speak out loud and voice overlap, and it also permits much systematic in-turn base environment of discussion for everyone.

The third in common opinion from the audience was the pronunciation of the language we employed for the experiment. Despite the fact that the English language is one that being spoken almost everywhere on the school campus, does not mean everyone is native, that being said, we have encountered several audiences who express an understanding issue, apparently caused by the speaker's way of speaking and pronouncing the language. Given the limited timeframe and resources, we could not involve a large group of native English or Finnish (concerning that the majority population on the campus are Finn) speakers to compare with non-natives. However, in the similar regard, we neither observed superior performance from the participants prepared a script in advance—assuming this way, contents are supposed to well organised as opposed to improvising.

Lastly, it was rather controversial results mapping the real speaker onto the virtual avatar in the display; five out of ten people accepted the concept of perceiving the two as the same person. One pointed out that the design of the avatar does not match with the look of the real speaker, thus, discourages the concept being digested.

6.4. Limitations

The study has several limitations. For example, taking a single variable to determine the performance quality as a whole may be insufficient. The interview data which was discarded from the analysis has a rich information about the impression of the system to the audience, though, it suffered from providing the assessment of speech quality. Similarly, another variable which is ‘spectators count’ has a relatively heavy dependency on the randomness of the current population of the region, and hence, the variable has to be exact equal across all speech sessions in order to build a robust assertion of the study.

While carrying out the experiments, we had to adjust a few experimental settings so we could collect any data at all. Although changing variables are flawed behaviour, we believe any of actions did not influence much on the result data in all aspects.

The small data set might have weakened the significance of the study. For example, Welch [84] suggested the impact of social factors in VE, which we also observed that some actions of the speaker other than talk had enticed more audience. However, the volume of data related to this was too small that we could not statistically prove the significance.

Lastly, we do not provide multifold analysis upon different variables introduced earlier, such as language selection, speech topic, display location, speech volume, CAVE illumination, date, and the timing; instead, we try to keep them consistent across all experiments.

7. CONCLUSION

We conclude the work in this thesis investigated the human's immersion and presence degree in relation to the task performance done in a VE. Our six successive experiments were conducted with the strong aim for evaluating the work fulfilment, and subsequently measuring the user's immersion and presence while being in VE.

We found that the speech duration has the effect on the general performance, and the tendency of favouring the interface of VW has positively associated with supplying a superior performance. However, we could not find a significant relation between general speech performance and the users' immersion and presence degree. Most of tests we ran on the relationship between human psychologic sensation and task fulfilment in a VE did not show a significance given the data sample.

Overall, our work has testified that the immersion and presence of a human utilising a VW for giving a speech, in particular, does not affect one another. To our best knowledge, this work has demonstrated a unique research subject and the approach to it.

8. BIBLIOGRAPHY

[1] (2016, Sep 16) Merriam-Webster. [Online]. http://www.merriam-webster.com/dictionary/soapbox

[2] G. A. Caldwell, M. Guaralda, J. Donovan, and M. Rittenbruch, "The InstaBooth: Making common ground for media architectural design", in *Media Architecture Biennale 2016*, Sydney, 2016.

[3] G. A. Caldwell and M. Foth, "DIY media architecture: open and participatory approaches to community engagement", In *Proceedings of the 2nd Media Architecture Biennale Conference: World Cities. ACM*, 2014, pp. 1-10.

[4] H. Jenkins, "Convergence culture: Where new and old media collide", *Nue-va York: New York UP*, 2006.

[5] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire", *Presence: Teleoperators and virtual environments*, vol. 7(3), pp. 225-240, Jun 1998.

[6] B. Mennecke et al., "Second Life and other virtual worlds: A roadmap for research", In *28th International Conference on Information Systems (ICIS)*, December. 2007.

[7] B. G. Witmer and M. J. Singer, "Measuring immersion in virtual environments", US Army Research Institute for the Behavioral and Social Sciences, Tech. Rep. 1014 1994.

[8] S. Johnstone, G. A. Caldwell, and M. Rittenbruch, "Defining the InstaBooth: Facilitating debate and content creation from situated users.", In *Proceedings: MEDIACITY 5 International Conference, Workshops and*

Urban Interventions, School of Architecture, Design & Environment and i-DAT (Institute of Digital Art and Technology), Plymouth, 2015, p. 187.

[9] R. Graham, *Anarchism: a documentary history of libertarian ideas*, 2nd ed.: Black Rose Books Ltd, 2009.

[10] I. Shklovski and M. F. Chang, "Urban computing: Navigating space and context", *COMPUTER-IEEE COMPUTER SOCIETY*, vol. 39, no. 9, p. 36, Sep 2006.

[11] T. Kindberg, M. Chalmers, and E. Paulos, "Urban computing", *IEEE Pervasive Computing*, vol. 6, no. 3, pp. 18-20, Jul 2007.

[12] Y. Liu. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs.", In *Proceedings of the 13th international conference on Ubiquitous computing. ACM*, 2011, pp. 89-98.

[13] N Davies, S. Slinch, and Alt Florian, *Pervasive displays: understanding the future of digital signage,* Synthesis Lectures on Mobile and Pervasive Computing, 2014.

[14] K. Galloway and S. Rabinowitz. (1980, July) Hole-In-Space. [Online]. http://www.ecafe.com/getty/HIS/

[15] G. O. Goodman and M. J. Abel, "Collaboration research in SCL.", In *Proceedings of the 1986 ACM Conference on Computer-supported Cooperative Work*, New York, 1986, pp. 246-251.

[16] S. A. Bly, S. R. Harrison, and S. Irwin, "Media spaces: bringing people together in a video, audio, and computing environment.", *Communications of the ACM*, vol. 36, no. 1, pp. 28-46, Jan 1993.

[17] R. S. Fish, R. E. Kraut, and B. L. Chalfonte, "The VideoWindow system in informal communication.", In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, 1990, pp. 1-11.

[18] G Jancke, G. D. Venolia, J. Grudin, J. J. Cadiz, and A. Gupta, "Linking public spaces: technical and social issues.", In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2001, pp. 530-537.

[19] K. Karahalios and J. Donath, "Telemurals: linking remote spaces with social catalysts.", In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 615-622.

[20] J. McCarthy, "Using public displays to create conversation opportunities.", In *Proceedings of Workshop on Public, Community, and Situated Displays at CSCW'02*, 2002.

[21] H. Brignull and Y. Rogers, "Enticing people to interact with large public displays in public spaces.", In *Proceedings of INTERACT*, vol. 3, 2003, pp. 17-24.

[22] T. Ojala et al., "UBI-hotspot 1.0: Large-scale long-term deployment of interactive public displays in a city center.", In *Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on*, 2010, pp. 285-294.

[23] T. Ojala et al., "Multipurpose interactive public displays in the wild: Three years later.", *Computer*, vol. 45, no. 5, pp. 42-49, 2012.

[24] A. Howard. (2005) Red Objects. [Online]. http://redobjects.unsw.edu.au/research/postgraduate/orchestrating-the-

public/

[25] Welch and Adrian. (2015) N Building Tokyo : Commercial Development Japan. [Online]. http://www.earchitect.co.uk/tokyo/n-building-tokyo

[26] T. Schneider and J. Till, "Beyond discourse: notes on spatial agency", *Footprint* , vol. 3, no. 1, pp. 97-112, Jan 2009.

[27] K. M. Kapp. (2007, April 07) Association for Talent Development. [Online]. https://www.td.org/Publications/Newsletters/Links/2007/04/Defining-and-Understanding-Virtual-Worlds

[28] T. Tuten, "Real world experience, virtual world environment: The design and execution of marketing plans in Second Life", *Marketing Education Review*, vol. 19(1), pp. 1-5, Mar 2009.

[29] N. Stephenson, *Snow Crash*. New York: Bantam Spectra, 1992.

[30] L. Carroll, *Through the Looking-Glass and What Alice Found There*. London: Macmillan and Co, 1883.

[31] R. Schroeder, "Defining virtual worlds and virtual environments", *Journal For Virtual Worlds Research*, vol. 1(1), 2008.

[32] A. S. Axelsson and T. Regan, *Playing Video Games: Motives, Responses, Consequences*, P Vorderer and J. Bryant, Eds.: Lawrence, 2006.

[33] E. Gobbetti and R. Scateni, "Virtual reality: Past, present, and future", *Virtual Environments in Clinical Psychology and Neuroscience: Methods and Techniques in Advanced Patient-Therapist Interaction*, 1998.

[34] J. M. Airey, J. H. Rohlf, and F. P. Brooks Jr, "Towards image realism with interactive update rates in complex virtual building environments", *ACM*

SIGGRAPH Computer Graphics, vol. 24, no. 2, pp. 41-50, Feb 1990.

[35] F. P. Brooks Jr, "Walkthrough—a dynamic graphics system for simulating virtual buildings", In *Proceedings of the 1986 workshop on Interactive 3D graphics*, Jan. 1987, pp. 9-21.

[36] J. F. BALAGUER and S. V. DE GENNARO, "A virtual reality project at CERN", *Computer Graphics*, vol. 30(4), pp. 40-48, Nov 1996.

[37] M. Arponen, "From 2D Base Map to 3D City Model.", In *Proceedings UMDS* , Prague, 2002.

[38] M. Carrozzino, C. Evangelista, and M. Bergamasco, "The Immersive Time-machine: A virtual exploration of the history of Livorno.", In *Proceedings of 3D-Arch Conference.*, 2009, pp. 198-201.

[39] J. Döllner, T. H. Kolbe, F. Liecke, T. Sgouros, and K. Teichmann, "The virtual 3d city model of berlin-managing, integrating, and communicating complex urban information", In *Proceedings of the 25th Urban Data Management Symposium UDMS*, 2006, pp. 15-17.

[40] T. Alatalo, T. Koskela, M. Pouke, P. Alavesa, and T. Ojala, "VirtualOulu: Collaborative, immersive and extensible 3D city model on the web", In *Proceedings of the 21st International Conference on Web3D Technology*, 2016, pp. 95-103.

[41] E. Palmquist and S. Jonathan, "Collaborative City Modeling", In *Proceedings Architecture in Computro* , 2008, pp. 249-256.

[42] M. L. Maher, M. Paulini, and P. Murty, "Scaling up: From individual design to collaborative design to collective design", *Design Computing and*

Cognition'10, 2011, pp. 581-599.

[43] J. P. Virtanen et al., "Intelligent Open Data 3D Maps in a Collaborative Virtual World", *ISPRS International Journal of Geo-Information*, vol. 4, no. 2, pp. 837-857, May 2015.

[44] M. Uden and A. Zipf, "Open building models: towards a platform for crowdsourcing virtual 3D cities", *Progress and New Trends in 3D Geoinformation Sciences*, 2013, pp. 299-314.

[45] F. Biljecki, J. Stoter, H. Ledoux, S. Zlatanova, and A. Çöltekin, "Applications of 3D city models: state of the art review," *ISPRS International Journal of Geo-Information*, vol. 4, no. 4, pp. 2842-2889, Dec 2015.

[46] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments.", *Presence: Teleoperators and virtual environments*, vol. 6, no. 6, pp. 603-616, Dec 1997.

[47] T. Alatalo, "An entity-component model for extensible virtual worlds", *IEEE Internet Computing*, vol. 15, no. 5, pp. 30-37, Sep 2011.

[48] unknown. (2013, October) soderhavet. [Online]. http://soderhavet.com/bl/2013/10/15/two-reasons-why-minecraft-is-succeeding-where-second-life-failed/#.V9Z-ofl96Uk

[49] A. De Lucia, R. Francese, I. Passero, and G. Tortora, "Development and evaluation of a virtual campus on Second Life: The case of SecondDMI", *Computers & Education*, vol. 52(1), pp. 220-233, Jan 2009.

[50] C. Youngblut, "Educational uses of virtual reality technology," Institute for defense analyses, Virginia, IDA D-2128, 1998.

[51] M. Roussos et al., "Learning and building together in an immersive virtual world", *Presence: Teleoperators and Virtual Environments*, vol. 8(3), pp. 247-263, Jun 1999.

[52] L. Jarmon, T. Traphagan, M. Mayrath, and A. Trivedi, "Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in Second Life", *Computers & Education*, vol. 53(1), pp. 169-182, Aug 2009.

[53] M. Honey, K. Connor, M. Veltman, D. Bodily, and S. Diener, "Teaching with Second Life®: Hemorrhage management as an example of a process for developing simulations for multiuser virtual environments", *Clinical Simulation in Nursing*, vol. 8(3), pp. e79-e85, Mar 2012.

[54] M. N. K. Boulos, L. Hetherington, and S. Wheeler, "Second Life: an overview of the potential of 3-D virtual worlds in medical and health education", *Health Information & Libraries Journal*, vol. 24(4), pp. 233-245, Dec 2007.

[55] K. Carr and R England, "Virtual Environments and Environmental Instruments", *Simulated and Virtual Realities*.: Taylor and Francis, 1996, pp. 85-101.

[56] J. Günther, *Digital natives & digital immigrants*. Innsbruck, Austria: Studienverlag, 2007.

[57] P. Hemp, "Avatar-based marketing", *Harvard business review*, vol. 84(6), pp. 48-57, Jun 2006.

[58] B. Brandherm, S. Ullrich, and H. Prendinger, "Simulation of sensor-based tracking in Second Life", In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: demo papers*, 2008, pp. 1689-1690.

[59] J. M. Rolfe and K. J. Staples, *Flight simulation*. Cambridge: Cambridge University Press, 1988.

[60] D. A. Bowman, "Virtual reality: how much immersion is enough?", *Computer*, vol. 40(7), pp. 36-43, Jul 2007.

[61] V. D. Veksler, "Second Life as a simulation environment: Rich, high-fidelity world, minus the hassles", In *Proceedings of the 9th International Conference of Cognitive Modeling*, 2009.

[62] M. C. Becker, P. Salvatore, and F. Zirpoli, "The impact of virtual simulation tools on problem-solving and new product development organization", *Research Policy*, vol. 34(9), pp. 1305-1321, Nov 2005.

[63] O. Kharif. (2006, October 17) Bloomberg. [Online]. http://www.bloomberg.com/news/articles/2006-10-17/big-media-gets-a-second-lifebusinessweek-business-news-stock-market-and-financial-advice

[64] J. Steuer, "Defining virtual reality: Dimensions determining telepresence", *Journal of communication*, vol. 42(4), pp. 73-93, 1992.

[65] M. Minsky, "Telepresence", *Omni*, pp. 45-51, June 1980.

[66] T. B. Sheridan, "Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments", *MIT Press*, vol. 1(1), pp. 120-

126, 1992.

[67] T. B. Sheridan and T. A. Furness, "Presence: Teleoperators and Virtual Environments", *MA: MIT Press*, vol. 1(1), 1992.

[68] R. M. Held and N. I. Durlach, "Telepresence. Presence: Teleoperators and Virtual Environments", *MIT Press*, vol. 1(1), pp. 102-112, 1992.

[69] J. J. Gibson, *The senses considered as perceptual systems*. Oxford, England: Boston: Houghton Mifflin, 1966.

[70] D. A. Norman, *The design of everyday things*. New York: Doubleday, 1988.

[71] D. A. Norman and S. W. Draper, *User-centered system design*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.

[72] K. Stanney and G. Salvendy, "Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda", *International Journal of Human-Computer Interaction*, vol. 10(2), pp. 135-187, Jun 1998.

[73] M. Slater, "A note on presence terminology", *Presence connect*, vol. 3(3), pp. 1-5, Jan 2003.

[74] D. Zeltzer, "Autonomy, interaction, and presence.", *Presence: Teleoperators & Virtual Environments*, vol. 1, no. 1, pp. 127-132, 1992.

[75] G. Fontaine, "The experience of a sense of presence in intercultural and international encounters", *Presence: Teleoperators & Virtual Environments*, vol. 1(4), pp. 482-490, 1992.

[76] A. M. Triesman, "Verbal cues, language, and meaning", *American Journal of Psychology*, vol. 77(2), pp. 206-219, Jun 1963.

[77] A. M. Triesman and J. G. A. Riley, "Is selective attention perception or selective response: A further test", *Journal of Experimental Psychology*, vol. 79(1p1), p. 27, Jan 1969.

[78] J. H. Bailey and B. G. Witmer, "Learning and transfer of spatial knowledge in a virtual environment", In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 38, no. 18, pp. 1158-1162, Oct 1994.

[79] M. J. Singer, J. Ehrlich, and S. Cinq-Mars, "Task Performance in Virtual Environments: Stereoscopic Versus Monoscopic Displays and Head-Coupling", ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA, ADA306720, 1995.

[80] A. Spagnolli and L. Gamberini. (2002) Immersion/Emersion: Presence in hybrid environments.

[81] V. Jr. Fernando, The Soapbox, Soapbox: real life image.

[82] Tundra Document Files. Document. [Online]. https://doc.meshmoon.com/doxygen/_tundra_documentfiles.html

[83] J. Dirksen, *Three. js Essentials*. Birmingham, UK: Packt Publishing Ltd, 2014.

[84] R. B. Welch, T. T. Blackmon, A. Liu, B. A. Mellers, and L. W. Stark, "The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence", *Presence: Teleoperators & Virtual Environments*, vol. 5(3), pp. 263-273, 1996.

[85] M. M. Wloka, "Interacting with virtual reality", *Virtual Prototyping*, 1995, pp. 199-212.

[86] C. A. Steinkuehler and D. Williams, "Where everybody knows your (screen) name: Online games as 'third places'", *Journal of Computer-Mediated Communication*, vol. 11(4), pp. 885-909, Jul 2006.

[87] L. B. Rosenberg and B. D. Adelstein, "Perceptual decomposition of virtual haptic surfaces", *Virtual Reality, 1993. Proceedings., IEEE 1993 Symposium on Research Frontiers in. IEEE*, 1993, pp. 46-53.

[88] M. Slater, M. Usoh, and A. Steed, "Depth of presence in virtual environments" *Presence: Teleoperators & Virtual Environments*, vol. 3(2), pp. 130-144, 1994.

[89] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory.", *Presence*, vol. 10, no. 3, pp. 282-297, Jun 2001.

9. APPENDICES

	Date	Begin time	Speech duration	Comments received	Votes	Language / Mother tongue	Active engagement	Passive engagement	Presentation method
Participant 1	Wed	12:29	11 mins	0	0	English / Norwegian	0	2	Reading from script
Participant 2	Thurs	10:06	18 mins	0	0	English / Russian	0	7	Improvised speech
Participant 3	Thurs	12:32	17 mins	2	1 like	English / Vietnamese	1	7	Improvised speech
Participant 4	Thurs	15:15	6 mins	0	0	English / Urdu	0	0	Improvised speech
Participant 5	Thurs	12:57	17 mins	0	0	English / Dutch	0	5	Reading from script
Participant 6	Thurs	14:23	23 mins	69	6 likes / 4	English / Chinese	11	16 or more	Improvised

Speech comprehensible scores in average
0
0
1
0
1
3.4

Appendix 1. Experiment report of the speakers (Audience who scanned the QR code falls into "Active engagement" group, and who did not but spent time watching/listening the speech falls into "Passive engagement" group).

Speakers	Audiences	Age	Gender	Background	Motivation
Participant 1	Audience 1	22	Male	Education	Curiosity
	Audience 2	23	Male	Education	Curiosity
Participant 2					
Participant 3	Audience 1	22	Female	Business	3D view was interesting.
	Audience 2		Male	Education	Unique system setting.
Participant 4					
Participant 5	Audience 1	25	Male	Computer science	I was near the hotspot.
Participant 6	Audience 1		Female	Education	After class found friends
	Audience 2		Male	Information processing	Knew a similar system before
	Audience 3	20	Female	Humanity	The speaker has a nice appearance. New technology.
	Audience 4		Male	Computer science	Curious as I have also the computer background.
	Audience 5		Female	Humanity	The speaker has a nice appearance.

Possession of the OR scanner	Regarding speech	General opinions	Virtual in a reality space
No	Couldn't hear the speech.	Too low sound volume.	Interesting
No	Couldn't hear the speech.	Too low sound volume.	Fine
No	Couldn't understand the pronounce.	Fine in general.	Avatar design prevents from focusing on the speech
No	Couldn't hear well, might be speaker's pronounce issue.	Looking good.	Interesting and I can imagine the avatar is the speaker.
No	Did not get the idea (speak fast). The topic is irrelevant to me.		I can imagine the speaker is represented by an avatar.
No	Did not listen.	System reminds of a live TV	Interesting
Yes	Like the activity with the speaker	Found some differences with which I used to know	Interesting
No	Can't get the idea, but was able to focus on the speech.	Voting function did not work in the display UI.	Confusing, not interesting
Yes	I was able to focus on the speech.	Mobile and web interface were new to me. It was	Interesting but don't feel the avatar is the speaker.
No	I agree to the topic very much and easy to focus on the speech	Sound was low and I would prefer direct chat instead of texting	Don't like the virtual part.

Appendix 2. Audience interview report.